

THIS WEEK

EDITORIALS

DISEASE Prevalence of diabetes soars in the United Arab Emirates **p.276**

WORLD VIEW Spanish science faces trouble and terminal decline **p.277**



JUST SO How the zebrafish got its stripes **p.279**

Worldwide weapons

Progress towards a United Nations arms-trade treaty is encouraging, but it won't keep weapons out of the hands of human-rights abusers.

Scientists are prominent among those trying to make the world a safer place. Albert Einstein was committed to the international peace effort, and Soviet nuclear physicist Andrei Sakharov and US chemist Linus Pauling are among the researchers who have been awarded the Nobel Peace Prize. The Polish-British nuclear physicist Joseph Rotblat received the peace prize in conjunction with the Pugwash Conferences on Science and World Affairs, the nuclear-disarmament organization that he helped to found (see *Nature* **481**, 438–439; 2012). The attitudes of these researchers chime with the internationalism of the scientific endeavour and the humanitarian goals that often motivate it.

At the same time, science and technology are integral to military development, and defence funding supports a great deal of research, much of it excellent. There need be no contradiction here: nations have a right to self-defence, and armed forces are often deployed for peace-keeping as well as for aggression. But what constitutes responsible use of military might is controversial, and peace-keeping is generally necessary only because aggressors have been supplied with military hardware in the first place.

All of this makes arms control a thorny subject for scientists. When, at a session on human rights at a physics conference several years ago, *Nature* asked whether the link between the arms trade and human-rights abuses might raise ethical concerns about research on offensive weaponry, the panellists shuffled in their seats and became tongue-tied.

There is no easy way to demarcate the ethical boundaries of defence research. But scientists should welcome progress towards a binding United Nations Arms Trade Treaty (ATT), for which a preparatory meeting in New York this week presages final negotiations in July.

The treaty aims to align the conditions and standards for arms exports from all signatory countries. The UK government's Foreign and Commonwealth Office (FCO), which supports the ATT, says that inconsistencies and loopholes in current regional and national control systems for the arms trade hinder sustainable development, undermine stability and democracy, and impede progress towards the UN's Millennium Development Goals.

Some nations will attempt to have the treaty watered down. That the sole vote against the principle at the UN General Assembly in October 2009 was from Zimbabwe speaks volumes about probable reasons for opposition. But let us not overlook the fact that in a vote a year earlier, Zimbabwe was joined by one other dissenter: the United States, then led by President George W. Bush. Would any of the current leading US Republican candidates for president be better disposed towards an ATT?

Even if it does go ahead, a treaty will not necessarily change the arms trade much. Most of the military technology used for human-rights abuses in recent decades has been obtained legally. Sales from Britain, for example, helped Libya's former leaders to suppress 'rebels' in 2011 and enabled Zimbabwe to launch assaults in the Democratic Republic of Congo in the 1990s.

➔ **NATURE.COM**
For more on science
and the military:
nature.com/military

The UK government admits that the ATT will probably not reduce arms sales. It says that the criteria for exports "would be based on existing obligations and commitments to prevent human rights abuse" — which have not been notably effective. According to the FCO, the treaty aims "to prevent weapons reaching terrorists, insurgents and

human rights abusers". But as demonstrated in Libya, one person's insurgents are another's democratizers, and today's legitimate rulers can be tomorrow's human-rights abusers.

"The Arms Trade Treaty could simply legitimize business as usual."

The FCO says that the treaty "will be good for business, both manufacturing and export sales". Indeed, arms manufacturers support it as a way of levelling the market playing field. The

ATT could simply legitimize business as usual by clearly demarcating it from the black market, and it will not cover peripheral military hardware such as surveillance and information technology. Some have argued that the treaty will be a distraction from the problem of keeping arms from human-rights violators (D. B. Kopel *et al. Penn State Law Rev.* **114**, 101–163; 2010).

So although there are good reasons to call for a strong ATT, it is no panacea. The real need is to establish what a responsible arms trade would look like, if this isn't an oxymoron. Some hard research is required on how existing, 'above-board' arms sales have affected governance, political stability and socio-economic conditions worldwide. This is challenging and contentious, but several starts have been made, both in the UN and nationally (see, for example, www.unidir.org and www.prio.no/nisat). We need more. ■

Tough choices

Scientists must find ways to make more efficient use of funds — or politicians may do it for them.

Scientists in the United States can find plenty of good news as they page through President Barack Obama's 2013 budget proposal. Despite substantial cuts elsewhere — and fierce pressure from Republicans to cut more — Obama called for healthy overall increases in both fundamental research and science education (see page 283).

But the good news, of course, is tempered by reality. Obama's budget document is one long struggle to balance two contradictory goals: to stimulate the lagging US economy and to curb the annual budget deficit, which is more than US\$1 trillion. Science and science education are widely viewed as helping with the first, and will doubtless continue to be seen as such no matter who wins November's presidential election.

The idea that science is a driver of prosperity is one of the few things on which the United States' bitterly divided political parties still agree. But the science funding agencies themselves are by no means immune to the second goal. The harder the cuts bite, the more those agencies will have to streamline their operations and merge or terminate programmes.

This week's budget proposal, which contains many references to "tough choices", shows that this process is already well under way. The Department of Energy (DOE), for example, wants to discontinue funding of several dozen projects that have not met their research milestones, or that seem otherwise unpromising. The National Science Foundation (NSF) is likewise cutting back on some \$66 million in lower-priority education, outreach and research programmes. The National Institutes of Health (NIH) has been ordered to pursue "new grant management policies" to increase the number of new grants by 7%. And NASA is being obliged to make drastic cuts to its Mars exploration programme so as to finish building its flagship James Webb Space Telescope.

Conceivably, this process could get even more drastic. Last month, Obama asked Congress to give him the authority to consolidate and streamline agencies on his own initiative — and suggested that one early application would be to transfer the National Oceanic and Atmospheric Administration from the Department of Commerce to the Department of the Interior. If Congress were to give Obama that power, it is possible to imagine him — or some future Republican president — sending all of the NSF's science-education programmes to the Department of Education, or merging the DOE's particle and nuclear physics research into the

NSF, under the guise of making management of science more efficient.

White House officials insist that no one in the administration is even contemplating such a wholesale restructuring. But the arithmetic of the deficit is unavoidable. Individual researchers, scientific societies and science funding agencies can no longer afford to be purely reactive, responding to each cut as it comes along. They need to be part of the debate, thinking systematically about how programmes and even whole agencies could be restructured to make them more efficient at using the scarce funds available, and more effective at promoting the best science.

To do that, and to address the increasing demands from politicians and voters for evidence that fundamental research is useful, scientists must also find better ways to measure the effectiveness of the nation's investments in science. The usual technique is to insist that principal investigators produce more and more reports, which tends to be a waste of everyone's time. A consortium of six universities called Star Metrics, launched in 2010 and headquartered at the NIH, has shown that it is possible to do better by using natural language processing and other tools to mine the data and reports that the agencies already collect. But even that is just a beginning. Researchers and research institutions need to help to devise still better measures — because if they don't do it themselves, politicians and others who know much less about science may very well do it for them. And who knows where that would end. ■

"Researchers, societies and funding agencies can no longer afford to be purely reactive."

On the up

The soaring incidence of diabetes is driving the United Arab Emirates' science ambitions.

In 1990, Dubai's desert skyline was flat. Literacy rates were low and the prevalence of diabetes was a modest 6%. The effects of oil-field wealth — both good and bad — had yet to kick in. Now the skyline soars with elegant skyscrapers, among them the world's tallest, the 830-metre-high Burj Khalifa. Literacy rates have risen with them to more than 90%, thanks to the enlightened policies of the ruling families in the United Arab Emirates (UAE). Unfortunately, as the residents of the federated country settled into sedentary, well-fed lifestyles, the prevalence of diabetes also soared — to well over 20%.

This is one reason why UAE science minister Sheikh Nahayan Mubarak Al Nahayan, who spoke to *Nature* earlier this month, sees merit in his country joining Europe's biobanking network, the Biobanking and Biomolecular Resources Research Infrastructure (BBMRI).

The BBMRI collects and shares standardized genetic and medical information on national populations. It is a long-term mega-project with a focus on complex diseases — including diabetes — that are caused by multiple genetic and environmental factors, and can be understood only by studying large numbers of people. Membership would support the UAE's focus on its principal medical problem, while building science capacity to international standards, one of its government's stated goals. For its part, the BBMRI is keen to add the emirates and its scientifically valuable population to the network.

Securing funding for the initiative won't be easy, however. The UAE comprises seven emirates, each ruled by its own royal family. Abu Dhabi is the largest by area — and, as home to most of the country's oil fields, the richest — whereas Dubai is the largest in terms of population. Between them, the royal families of these two emirates share the bulk of power in the federal government, which has a limited budget, mostly contributed by Abu Dhabi.

This complex political constellation has hindered the UAE's ambition to become a scientific force. For example, in 2008 the federal

government created the National Research Foundation (NRF) to support and encourage competitive research. Its annual budget was set at 100 million UAE dirham (US\$27 million) and it opened calls for centres of excellence and for individual investigators. Winners, identified by international peer review, were congratulated with great fanfare, but most of the promised funds for the NRF never arrived. The culture of science is new and foreign to UAE rulers, who are used to buying in whatever they need.

The process did at least provide the NRF with an effective audit of the small UAE research base. This has helped it to lobby successfully for money from industry, foundations and states on a case-by-case basis for some of the winners whose projects most visibly align with national interests, such as research on water resources. This is not how an independent research agency should operate, but positive feedback from the funded projects may eventually encourage proper funding of the NRF.

The NRF exercise also identified genetics and disease as a key research strength in the emirates, and similar lobbying is likely to see a centre of excellence established in this field. This should give hope to UAE scientists interested in participating in the BBMRI.

Those scientists met their European colleagues for first discussions earlier this month in Dubai, where they will also hold a scientific symposium in October. A formal proposal for funding will emerge from the symposium, and the UAE government would be wise to support it. UAE researchers say that the state has three natural resources: oil and gas, dates and camel milk. In decades to come, the oil and gas will run out. Membership of the BBMRI could provide a short cut in the long process of building a knowledge-based economy.

The other Gulf states share the UAE's predisposition to diabetes, as well as its dependence on oil and gas reserves. They could also benefit from joining the project, and extending the Middle Eastern population base would make the research more powerful.

Will that happen? The Gulf states tend to be very competitive. Unnerved by the Burj Khalifa, Saudi Arabia planned to build its Kingdom Tower to be 1,000 metres. And Kuwait's Mubarak al-Kabir Tower is proposed to be 1,001 metres. The BBMRI offers an alternative model, an umbrella under which participants cooperate and share. This is a different mind set, but one that is necessary to resolve the region's shared health problems. ■

➤ NATURE.COM
To comment online,
click on Editorials at:
go.nature.com/xhunjv



Spanish changes are scientific suicide

If research continues to be sidelined, Spain will be left with little domestic expertise, warns Amaya Moro-Martín.

Spain no longer has a ministry of science. In the last days of 2011, its new government transferred national science policy to the Ministry of Economy and Competitiveness, a duty for which this ministry seems most unsuited. Science was an unwelcome addition that absorbed more than half of the €1,083-million (US\$1,438-million) budget cut imposed on the ministry. This sends an alarming signal of the sacrifices that science may face when the government releases its budget for 2012 next month.

This is the first time that neither 'science' nor 'research' have featured in the name of any top Spanish government department. It is not just a symbolic shift: it continues our country's trend of deliberately undermining and playing down the importance of science.

The official line is clear: science is not a priority in Spain. Of course, we are immersed in an economic crisis and austerity measures are needed. However, the government's irrational and draconian actions will cause long-term damage to the scientific infrastructure and send contradictory messages to other countries and investors. Although its rhetoric promises a shift to a knowledge-based economy, every step it takes is in the opposite direction. The results will be a borrowed-knowledge economy with little domestic know-how.

The problems did not start with the new government: the previous administration attempted to pass a Kafkaesque by-law for public universities that would have created a merit-evaluation system that diminished the weight assigned to research and technology transfer. The by-law stated that trade unions would negotiate the criteria for faculty promotion, making academic careers "more predictable and more egalitarian". It would have been the death of meritocracy. The same by-law would also have ballooned bureaucracy to such a level that it would have threatened to swamp any university administration.

The previous government also opposed to create a genuine tenure-track system for researchers in universities and national laboratories, on the grounds that tenure track is unconstitutional because access to civil service should be "egalitarian" so tenured jobs should not be targeted to tenure-track researchers. This is a consequence of the narrow-minded thought that all researchers in the public sector should be civil servants, but civil service is unsuited to research activities.

Spain likes to boast that it has an equivalent to tenure track: the Ramón y Cajal programme. Launched in 2001, this is the only nationwide programme that has managed to attract and retain highly qualified researchers from Spain and abroad. However, drastic cuts in hiring over the past three years and a hiring freeze announced this year will kill this first

attempt at a tenure-track programme. The prospects are so grim that despite being eager to return to Spain, some of my Spanish colleagues in the United States are rejecting Ramón y Cajal positions.

The hiring freeze is suicidal. Researchers who retire will no longer be replaced. Unlike many of its neighbours, Spain has a very limited science and technology industry in which to absorb highly qualified workers, so scientists aged 20–40 years will have no choice but to leave if they want to further their career. The country will therefore face a multigenerational brain drain, with corresponding losses in innovation, inspiration and credibility. The damage from this decision will take decades to reverse.

The new government is now effectively trampling on the best hope that Spanish researchers had for the future. Legislation in the pipeline could have improved the situation, but the government has, abruptly and without explanation, closed the two political science commissions — one in the Senate and one in the Congress — that would have been responsible for steering through this legislation.

The legislation includes moves to allow universities and research centres to be funded privately, to develop a new science and technology strategy and to create a proper national research agency with a multi-year budget. We urgently need such a system in Spain, where severe and unpredictable fluctuations in year-to-year funding make medium- to long-term planning impossible. The strategy is crucial if Spain is to coordinate its increasingly anarchic 18 sets of science policies — laid out simultaneously by the 17 regional governments and the central government — and to introduce a smarter, top-down, approach to tackling national problems.

Spain must bring its science and technology investment (currently 1.39% of gross domestic product) in line with European standards (2%) and closer to the 3% goal set by the European Council Lisbon Strategy for 2010. It also needs a science council, similar to the German *Wissenschaftsrat*, constituted mainly of scientists who have been elected by the scientific community to take the lead in delivering the national science and technology strategy.

Spain's situation is summed up by a poster for a recent Hollywood blockbuster: "No plan. No backup. No choice. Mission: Impossible. Ghost Protocol." Spanish science cannot afford ghost protocols. Without the proposed strategy there is no plan, and without a well-funded and non-political national research funding agency, there is no backup. The results leave research in Spain with a mission impossible. ■

Amaya Moro-Martín is a Ramón y Cajal Fellow at the Spanish National Research Council in Madrid.
e-mail: amaya@cab.inta-csic.es

**THE GOVERNMENT'S
IRRATIONAL
AND DRACONIAN
ACTIONS WILL CAUSE
LONG-TERM
DAMAGE
TO THE SCIENTIFIC
INFRASTRUCTURE.**

➔ **NATURE.COM**
Discuss this article
online at:
go.nature.com/cgja3

RESEARCH HIGHLIGHTS

Selections from the
scientific literature

CLIMATE MODELLING

A rainy signal from noise

A temperature increase of at least 1.4 °C is needed before changes in regional precipitation can be distinguished from regular variability and attributed to global warming.

A team led by Irina Mahlstein of the National Oceanic and Atmospheric Administration in Boulder, Colorado, used a suite of general circulation models to analyse regional precipitation trends from 1900 to 2099. The analysis focused on wet seasons, for which models performed most accurately against historical trends.

By the end of this century, the study suggests, increases in wet-season precipitation will be apparent in many areas. However, the authors note that changes in extreme weather or annual precipitation might be detectable much earlier.

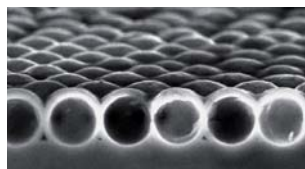
Geophys. Res. Lett. <http://dx.doi.org/10.1029/2011GL050738> (2012)

MATERIALS

Nanoscale shells trap light

Sheets of silicon nanoshells created by a team in California could lead to ultra-thin solar panels that are cheaper and easier to mass-produce than those currently available.

Conventional solar panels absorb light using relatively thick layers of nanocrystalline silicon that



can be time-consuming to manufacture. Yi Cui and his colleagues at Stanford University manufactured spherical, hollow silicon shells using standard chemical techniques and deposited them on a sheet (pictured). Light captured by the material was reflected many times inside the shells, increasing the amount of energy the sheet absorbed.

The team found that a 50-nanometre-thick layer of shells was as efficient as a 1-micrometre-thick sheet of conventional silicon.

Nature Commun. 3, 664 (2012)



MICROBIOLOGY

Seal corpses shelter Antarctic microbes

Mummified seals scattered across the deserts of Antarctica's McMurdo Dry Valleys reveal that microbial communities in the region respond rapidly to environmental change.

The seal carcasses are naturally mummified by the extremely dry, cold conditions of one of the world's least hospitable climates. Craig Cary at the University of Waikato in Hamilton, New Zealand, and his colleagues found that undisturbed carcasses boost humidity, stabilize temperature and alter the microbial

communities in the soils beneath them.

The researchers assayed how quickly those communities changed by transplanting a 250-year-old seal carcass to a pristine location. Two summers later, the microbial composition of the soil beneath the transplanted seal resembled that of the seal's original location. This challenges the hypothesis that the region's soil ecology changes only over the course of centuries.

Nature Commun. 3, 660 (2012)

BEHAVIOURAL SCIENCE

Sex is spread across the genes

Sex-specific behaviours in activities such as mating and parenting are controlled in a modular way by distinct sets of genes.

Nirao Shah at the University of California, San Francisco, and his colleagues screened the brains of male and female mice for differences in gene expression. They identified 16 genes differentially expressed in the hypothalamus

and amygdala, brain regions implicated in the control of sex-related behaviours. Sex hormones, which drive behavioural differences between the sexes, exert their effects by regulating the expression patterns of these genes.

Mice deficient in one of the genes demonstrated subtle differences in particular sex-specific behaviours, such as female acceptance of — or male interest in — penetration, without affecting other sex-typical behaviours. *Cell* 148, 596–607 (2012)

ECOLOGY

Fish figures hint at past extinctions

Contrary to the popular saying, there are not plenty of fish in the sea. But why? Perhaps because a huge number of species became extinct in ancient times, say Greta Carrete Vega and John Wiens at Stony Brook University in New York.

Marine environments cover about 70% of Earth's surface but contain only 15–25% of all estimated species. To find out why, Vega and Wiens studied actinopterygian (ray-finned) fish — which encompass 96% of Earth's fish species — in marine and freshwater environments.

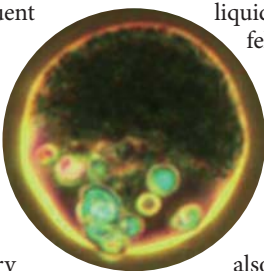
They found that both environments were similarly rich in actinopterygian species, even through the marine environment is much larger and has greater primary productivity. They also discovered that all extant marine actinopterygians descend from a freshwater ancestor, suggesting that ancient extinctions have robbed the seas of their species. *Proc. R. Soc. B* <http://dx.doi.org/10.1098/rspb.2012.0075> (2012)

CLIMATE CHANGE

More super-hot summers ahead

Summer temperatures once considered exceptionally high have, in recent decades, become more frequent across the United States as a result of anthropogenic climate change.

Philip Duffy, currently at the Lawrence Livermore National Laboratory in Livermore, California, and Claudia Tebaldi at the National Center for Atmospheric Research in Boulder, Colorado, compared summer temperature extremes from 1950 to 1999



with simulations derived from 16 global climate models. Model projections suggest that US summer temperatures will continue to rise as the century progresses. Even in regions that have warmed relatively little so far, the chances of extreme temperatures — seen only once in 20 years in the second half of the past century — will be at least 70% in any given year by 2064.

Clim. Change <http://dx.doi.org/10.1007/s10584-012-0396-6> (2012)

MATERIALS SCIENCE

Six-faced particles

Janus particles, named after the two-faced Roman god, are solid particles of two halves, each with different physical properties and with applications that include drug delivery. But Shoji Takeuchi and his colleagues at the University of Tokyo have gone beyond Janus's two faces and made gel spheres with up to six distinct sections. The particles are about 100 micrometres in diameter and, when each section is permeated with different fluorescent nanobeads, look like beach balls.

To make them, the researchers injected dyed sodium alginate solutions down a multi-barrelled capillary tube, which they then centrifuged. This forced the liquid streams out to form droplets composed of multiple liquids. The droplets fell into a waiting bath of calcium chloride, turning them to gel before the component solutions could mix.

The researchers also used their method to produce Janus particles holding magnetic particles and living cells in their separate halves (**pictured**). *Adv. Mater.* <http://dx.doi.org/10.1002/adma.201102560> (2012)

COMMUNITY CHOICE

The most viewed papers in science

IMMUNOLOGY

T-cell retreat in chronic hepatitis C

HIGHLY READ
on www.rupress.org in January

After acute infection with hepatitis C, some people recover whereas others develop chronic disease. Contrary to expectation, the latter group launches the same initial immune response against the virus as the former.

Recovery from hepatitis C infection was thought to be heralded by a broad response by CD4⁺ T cells, which had not previously been detected in most patients with chronic disease. But Georg Lauer of Massachusetts General Hospital in Boston and his colleagues detected these cells in the blood samples of 31 patients with acute infection, including 13 who later advanced to chronic disease. The researchers cultured the immune cells and, with the aid of sensitive fluorescent labelling, showed that the cells later disappeared from samples taken from the chronic disease group.

Early antiviral therapy can prevent the loss of CD4⁺ T cells, suggesting a possible means by which to prevent the development of chronic infection.

J. Exp. Med. 209, 61–75 (2012)



ZOOLOGY

Stripes from shifting cells

Repulsion between pigment cells helps to explain how adult zebrafish develop the stripes for which they are named.

Shigeru Kondo and his colleagues at Osaka University in Japan looked at cultured black melanophore and yellow xanthophore pigment cells from the animals (**pictured top**). They found that when the black pigment cells came into contact with the yellow ones, their membrane potential changed, shifting

the charge at the cells' surface. In 60% of such encounters, the melanophores moved away from the xanthophores. However, in mutant fish that lack regular stripes (**bottom**), the two cell types always stayed in contact with each other.

The authors say that although repulsion alone is not sufficient to explain pattern formation, repulsion could be involved in the stripes' development.

Science 335, 677 (2012)

► **NATURE.COM**

For the latest research published by Nature visit:
www.nature.com/latestresearch

SEVEN DAYS

The news in brief

POLICY

US budget hopes

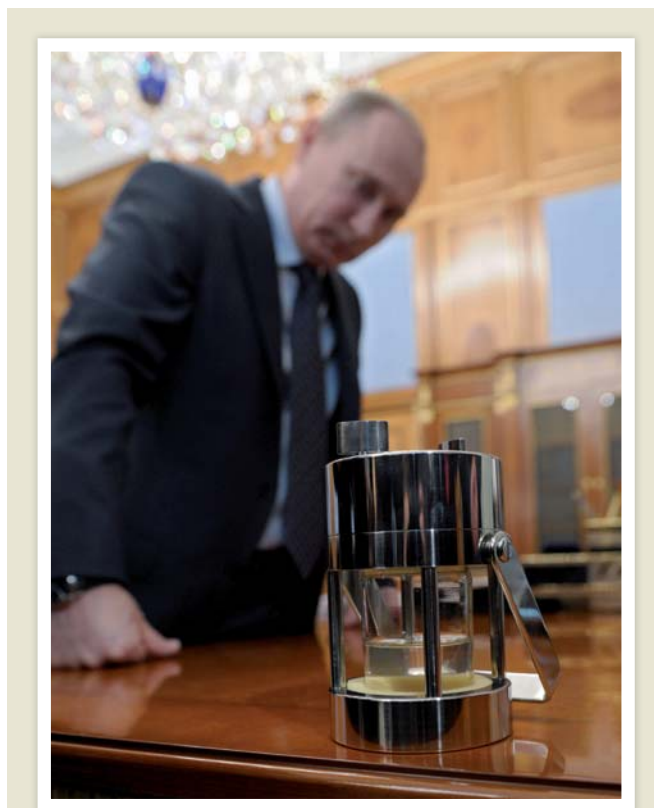
US President Barack Obama proposed slight increases for most major science agencies in his 2013 budget request, released on 13 February. The country's largest research agency, the National Institutes of Health, saw its budget held level. NASA, meanwhile, looks set to lose out, with cuts of 3.2% to its science budget and 21% to planetary science — leading the agency's administrator Charles Bolden to cancel plans for joint Mars missions with the European Space Agency. However, the forthcoming presidential election and existing agreements to trim government spending mean that next year's science budget remains very uncertain. See pages 283–285 for more.

Elsevier boycott

The number of researchers who have signed a public pledge not to support academic-publishing giant Elsevier, headquartered in Amsterdam, passed 5,000 last week. Mathematician Timothy Gowers at the University of Cambridge, UK, began the boycott with a 21 January blog opposing the company's practices, which he says hinder the dissemination of research. By 14 February, 5,847 had signed an online petition. A similar campaign in 2000–01, which attracted 30,000 signatories, was connected to the founding of the Public Library of Science publishing venture headquartered in San Francisco, California. See go.nature.com/uzkmay for more.

Nuclear clean-up

The Japanese government has threatened to withhold about ¥1 trillion (US\$12.8 billion) in rescue funds for the private



Putin's subglacial sample

Russia's Arctic and Antarctic Research Institute in St Petersburg confirmed on 8 February that scientists have managed to drill 3,769 metres through Antarctica's ice sheet to reach the subglacial Lake Vostok. The breakthrough was made on 5 February, the institute said. By 10 February, Russia's prime minister, Vladimir Putin, had acclaimed the discovery, and was presented on national television with a sample of yellowing water (pictured) — although the water was probably from melted ice at the bottom of the borehole, not from the lake itself. The Russian drilling team has now left the borehole until next summer (in December), when they will return to do further analysis. See page 287 for more.

company that runs the stricken Fukushima Daiichi nuclear power plant, unless it gets more say in the firm's operations. On 13 February, Japan's energy and trade minister, Yukio Edano, did approve ¥690 billion for the Tokyo Electric Power Company (TEPCO), which is currently struggling to pay compensation costs and clean

up the plant after it was hit by a tsunami last year. But, Edano said, the larger separate bailout would depend on TEPCO ceding partial control to the government.

US nuclear approval

The United States has given the green light to its first new nuclear reactors since 1978. On 9 February, the

US Nuclear Regulatory Commission (NRC) approved an application by utility giant Southern Company, based in Atlanta, Georgia, to build two pressurized-water nuclear reactors at its Vogtle station near Waynesboro. The company said the reactors could be operating by 2017. However, a US nuclear renaissance seems unlikely, because few other reactor proposals are in the pipeline. See go.nature.com/tws1oz for more.

RIA NOVOSTI

BUSINESS

Biosimilars rules

Drug-makers keen to sell generic forms of branded biological drugs — such as enzymes and antibodies — were excited to finally see draft guidance on the matter emerge from the US Food and Drug Administration (FDA) on 9 February. Proteins are large and complex, so it is much harder to copy drugs based on them than small-molecule drugs (see *Nature* **449**, 274–276; 2007). The FDA wants firms to prove their molecules' similarity to branded biologics before the generics can be approved — but the agency provided few concrete details, instead saying that it would judge on a case-by-case basis. See go.nature.com/nhbvik for more.

Illumina takeover

Illumina, the dominant developer of DNA-sequencing technology, has, as expected, rejected a US\$5.7-billion takeover bid by drug giant Roche, based in Basel, Switzerland. On 7 February, the board of directors of Illumina, which is headquartered in San Diego, California, said that the 25 January offer was “grossly inadequate”, undervaluing its

company's prospects. Roche replied that its bid was "full and fair". The firm is now expected to start wooing Illumina's major investors to accept a takeover — even though Illumina's share price is currently well above Roche's bid.

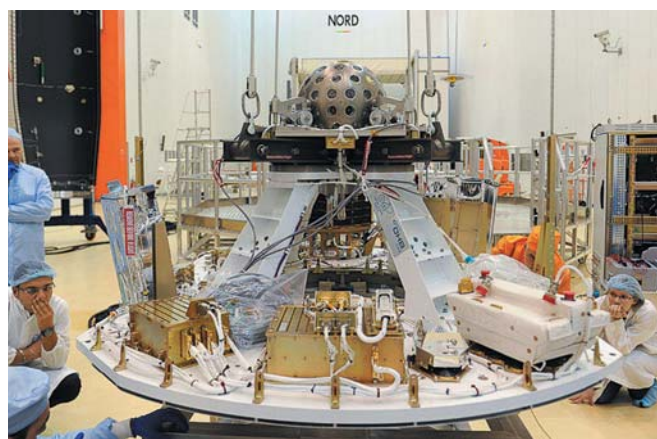
Synbio troubles

US synthetic-biology firm Amyris — which engineers microbes to process plant sugars into useful chemicals — saw its share price plunge by 28% on 10 February, after it admitted that it would not meet its production forecasts for a key chemical, farnesene. The firm, based in Emeryville, California, also said that it would not use its own microbe vats to produce synthetic fuels, leaving high-volume efforts to oil company Total, based in Paris, and biofuel firm Cosan in São Paulo, Brazil, with whom it has signed joint ventures.

RESEARCH

Vega launches

Europe's Vega rocket, a low-cost launcher intended to get small scientific satellites into low-Earth orbit, had a successful maiden flight on 13 February. The inaugural launch, from the European Space Agency's spaceport in Kourou, French Guiana, carried nine satellites; its main research payload was



the Italian Space Agency's Laser Relativity Satellite (LARES, pictured: sphere on top of the rocket's payload) which will study the Lense–Thirring effect, a distortion of space-time caused by Earth's gravity. The Vega rocket has cost more than €700 million (US\$924 million) to develop; five further flights are planned before 2016. See go.nature.com/srl2fb for more.

LHC schedule

On 13 February, operators of the world's most powerful particle accelerator announced their plan for its 2012 run, which starts in March. The new schedule calls for the Large Hadron Collider (LHC), located near Geneva, Switzerland, to smash protons together at an energy of 8 teraelectronvolts (TeV), an increase of 1 TeV

over the previous year, but still well short of the 14-TeV collision energy that the reactor was originally designed to reach. The team expects the LHC to produce around 1,600 trillion proton–proton collisions this year, a threefold increase over 2011. See go.nature.com/xivplh for more.

Denisovan genome

The complete genetic sequence of an extinct relative of humans — the Denisovan — was posted online (see go.nature.com/vvtcfi) on 6 February, allowing others to download the data while the work awaits formal journal publication. Researchers from the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany, mapped every position in the genome an average of

COMING UP

16–17 FEBRUARY

In Geneva, Switzerland, the World Health Organization will gather experts to discuss 'urgent questions' about research censorship and public safety, relating to unpublished work on mutant, transmissible strains of the H5N1 influenza A virus. See page 289 for more on the flu-virus debate.

go.nature.com/pf7bhw

20–24 FEBRUARY

Marine scientists' responses to the Gulf of Mexico oil spill in 2010 are among topics discussed at the Ocean Sciences Meeting in Salt Lake City, Utah.

www.sgmeet.com/osm2012

30 times, improving on the 1.9-fold coverage in their 2010 draft genome (D. Reich *et al. Nature* **468**, 1053–1060; 2010). A 30,000–50,000-year-old finger bone found in the Denisova Cave, southern Siberia, in 2008 yielded the genetic material. See go.nature.com/w3evow for more.

PEOPLE

China science prize

Chinese physicist Xie Jialin, who pioneered the building of China's first high-energy linear particle accelerator in 1964, and helped to develop its first free-electron laser in 1993, has won his nation's top science and technology award, worth 5 million renminbi (US\$794,000). The prize — which has been awarded by the country's president annually since 2000 — was presented on 14 February. Architect and town planner Wu Liangyong also won.

► NATURE.COM

For daily news updates see:

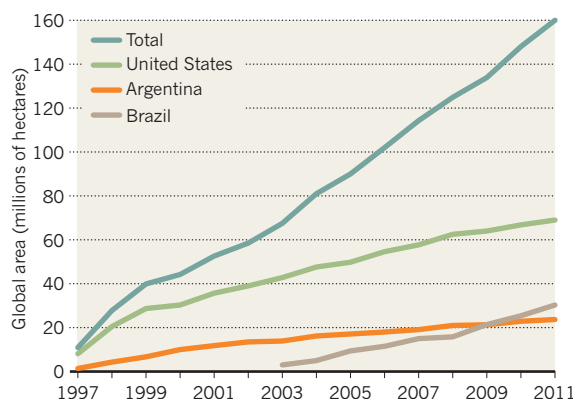
www.nature.com/news

TREND WATCH

Brazil has continued its rapid rise in planting of commercial genetically modified (GM) crops. The country, which is the world's second-largest adopter of such crops, grew 30.3 million hectares of GM soya, maize (corn) and cotton last year, a 19% increase on 2010. Argentina — which plants similar crops and is the third-largest adopter — crept up 3% to 23.7 million hectares. Both stay well behind the United States, which planted 69 million hectares in 2011. In total, 29 countries now plant GM crops.

BRAZIL DRIVES GM CROP GROWTH

Planting of genetically modified (GM) crops grew by 8% in 2011 to 160 million hectares worldwide.



NEWS IN FOCUS

ANTARCTICA Drilling success opens the way to exploring entombed lake **p.287**

GENOMICS Sequencing set to become a standard clinical test **p.288**

BIRD FLU How deadly is H5N1? Uncounted cases spark debate **p.289**



HUMAN EVOLUTION Did climate change drive ancient cultural leaps? **p.290**

S. LOEB/AFP/GETTY



Barack Obama aims high with an air cannon at this year's science fair at the White House.

US BUDGET

Obama shoots for science increase

US president wants to make room for research to grow in 2013 — but faces an uphill battle.

BY IVAN SEMENIUK, MEREDITH WADMAN, SUSAN YOUNG, ERIC HAND, EUGENIE SAMUEL REICH & RICHARD MONASTERSKY

“It’s not every day you have robots running through your house,” Barack Obama quipped last week at the White House science fair, a showcase for student exhibitors that also gave the US president a chance to reiterate a favourite theme. Science and technology, he said, “is what’s going to make a difference in this country, over the long haul”.

Obama would clearly like to see many more robots, as well as researchers and engineers, running around in the future, a wish reflected in his budget request for fiscal year 2013, released on 13 February. The document’s message is one of big ambitions with fewer resources.

A year ago, Obama proposed bold increases for science agencies, but a Congress intent on curbing government spending refused to back many of them. This time, the White House has scaled back in several areas but boosted overall funding for non-defence research and development by 5%, pushing it up to US\$64.9 billion.

“Overall, the budget sustains an upward trend,” says John Holdren, director of the White House Office of Science and Technology Policy in Washington DC. “Because of fiscal restraints, it’s not at the rate we preferred.”

With an election coming this November, House Republicans are unlikely to be generous with the president’s request. As in previous years, Congress could delay action on the budget, especially if it decides to wait for voters to weigh in on Obama’s presidency before making its decision. And the spectre of a severe across-the-board cut dangles over the government because of an act introduced last year that aims to chop \$1.2 trillion from spending, starting in January 2013.

Here is an overview of what the president’s request would mean in key science domains (see “Tough decisions”).

BIOMEDICINE AND PUBLIC HEALTH

The National Institutes of Health (NIH) in Bethesda, Maryland, by far the largest US research agency, sees its budget held level at \$30.7 billion — a far cry from the \$1-billion increase Obama proposed a year ago. Despite the ceiling, Lawrence Tabak, the NIH’s ▶

► principal deputy director, sees the budget as “continuing our priorities in basic science”, and it allows the agency to boost the number of new and competing grants it funds by 8%.

The newly launched National Center for Advancing Translational Sciences (NCATS) in Bethesda will grow by 11%, to \$639 million. Much of the rise goes to the Cures Acceleration Network, an effort to spur development of badly needed medicines through bold, multi-million-dollar grants. The programme's allocation grows fivefold next year, to \$50 million.

Accomplishing all this within a flat budget requires cuts. Losers include the National Children's Study, a long-term study of early influences on the health of more than a hundred thousand children, which received \$194 million in 2012, but has been cut by \$29 million; and the Institutional Development Award programme, aimed at developing research infrastructure in rural and underserved states, which loses nearly \$48 million.

To pinch the pennies that will make new grants possible, the NIH plans to eliminate inflationary increases for some existing grants, cut others by 1% and keep grants seeking renewal at current levels. The agency predicts that these measures would boost the success rate for grant applications, currently at a historic low of 18%, but only to 19%.

The flat-lined budget has drawn bleak appraisals from NIH advocates. “We are talking about a budget that is probably close to 20% smaller than it was a decade ago, adjusted for inflation,” says David Moore, senior director for government relations at the Association of American Medical Colleges in Washington DC.

Jennifer Zeitzer, director of legislative relations at the Federation of American Societies for Experimental Biology in Bethesda, says that her organization will work with research champions to persuade Congress to boost the allocation for the NIH. The president's request, she says, “is not what we need to take advantage of the scientific opportunities that are before us”.

The outlook is even less favourable for the Centers for Disease Control and Prevention (CDC) in Atlanta, Georgia, which has had its budget cut by 12%, to make a total of 22% in cuts since 2010. Those cuts are, in part, counterbalanced by bringing in funds from a long-standing health-services evaluation fund and from the Prevention and Public Health Fund, which is part of the health reform law that Obama introduced in 2010.

The dependence on the Prevention and Public Health Fund worries public-health advocates. Using the fund to patch holes in the CDC's budget is “troubling”, says Emily Holubowich, executive director of the Coalition for Health Funding, based in Washington DC. “The future of the fund is tenuous at best.”

Obama has also kept the budget mostly flat for the Food and Drug Administration (FDA) in Silver Spring, Maryland. However, the

TOUGH DECISIONS

The winners and losers in President Barack Obama's budget request for 2013 (US\$ millions).

Agency	2011 actual	2012 estimated	2013 requested	Details
Biomedical research and public health				
National Institutes of Health	30,470	30,702	30,702	Flat funding overall but 11% boost for translational science centre
Centers for Disease Control and Prevention	5,726	5,732	5,068	Efforts in public health and disease prevention bear the brunt of a deep cut
Food and Drug Administration	2,403	2,506	2,517	Flat government funding, but surge in industry user fees lifts overall budget by 17% to \$4,486 million
Physical sciences				
National Science Foundation	6,806	7,032	7,372	Big gains for interdisciplinary ideas — as well as marketable ones
NASA (science)	4,919	5,074	4,911	Flagship telescope still on track, but future of Mars exploration less certain
Department of Energy Office of Science	4,897	4,874	4,992	Overall increase masks cuts to some basic-research programmes
National Institute of Standards and Technology	754	761	860	Substantial increase, with around half to advanced manufacturing research
Earth and environment				
Environmental Protection Agency	8,681	8,450	8,344	Despite ongoing declines, core science and regulatory programmes preserved
National Oceanic and Atmospheric Administration	4,727	5,014	5,179	Modest rise, with satellite programme getting much-needed boost
US Geological Survey	1,084	1,068	1,102	Increased money for disaster response and research on hydraulic fracturing

Source: White House Office of Management and Budget

agency will receive a \$583-million bolus from new industries, mainly from food-registration and inspection fees and from makers of generic drugs and biosimilars.

The FDA has already been criticized for becoming too reliant on industry funding, but Margaret Hamburg, the FDA's commissioner, says that the fees are needed to ensure effective and timely drug and device review. “There is a common good here,” she says.

PHYSICAL SCIENCES

The White House continues to support a long-term doubling of budgets for physical-science agencies, including the National Science Foundation (NSF) in Arlington, Virginia; the Department of Energy's Office of Science in Washington DC; and the National Institute of Standards and Technology (NIST) in Gaithersburg, Maryland. The doubling, relative to 2006, is a goal of the America COMPETES Act, introduced under former President George W. Bush that year, and signed into law in 2007. Congressional appropriators have, however, slowed the pace of these agencies' growth considerably since then (see ‘A long way to go’).

The budget also shifts funding towards the applied end of the research spectrum, where advances should translate into economic gains more quickly. It continues to fund I-Corps, a programme launched last year that partners entrepreneurs with scientists seeking to test the marketability of their research. And advanced manufacturing, which supports industry by

developing measurement capabilities and standards to guide new product development, gets \$149 million — money that NSF director Subra Suresh says will help to stem a decline in US manufacturing. “In times of constrained budgets, we need to be crystal clear about why NSF matters,” Suresh says.

The NSF emerges as a clear winner in Obama's request, with a 5% boost to its bottom line. And one thing is very clear at the agency: researchers pursuing interdisciplinary research will be rewarded, with \$63 million allocated to a programme that supports such work.

The NIST also gets a large increase, much of which is aimed at advanced manufacturing, including both a robotics programme and a ‘materials genome’ initiative that aims to speed up the development of new materials.

The Department of Energy's Office of Science receives a more modest rise, much of which goes to its national laboratories and Energy Innovation Hubs. Several basic-research programmes are trimmed, including nuclear physics and high-energy physics, a shift that is consistent with the administration's emphasis on applied research that is most relevant to energy technology.

“Basic research is systematically down,” says Milind Diwan, a physicist at Brookhaven National Laboratory in Upton, New York, and co-spokesman for a planned particle physics experiment that received a drop in funding. “Those of us in fundamental-research have to live within those priorities.”

At NASA, the talk is of “tough but sustainable choices” for an agency that would receive \$17.7 billion in 2013, \$59 million less than in 2012. Its science budget drops by 3.2%, but planetary science bears the brunt of that, with a cut of 21%. For years, NASA has been pursuing plans with the European Space Agency for joint missions to Mars in 2016 and 2018. But on Monday, NASA administrator Charles Bolden pulled the plug. “We just cannot do another flagship right now,” he said. Officials fear that the costs for these missions would spiral out of control, as they have for the \$8.8-billion James Webb Space Telescope, a follow-on to the Hubble telescope that is slated for a 2018 launch.

The pinch will perhaps be felt most keenly at the Jet Propulsion Laboratory (JPL) in Pasadena, California, the traditional home of the Mars Exploration Program. Last year, the laboratory had to lay off the equivalent of 246 full-time employees, reducing its staff to 5,047. When the \$2.5-billion Mars Science Laboratory lands in August, the JPL will have to quickly find new work for a few hundred employees so the latest Mars cancellations make more lay-offs likely. “Our expectation was that we’d have another mission to move these people on to,” says Richard O’Toole, the JPL’s manager of legislative affairs. “We definitely feel the pressure.”

ENERGY, EARTH AND ENVIRONMENT

In Obama’s plan, spending on energy efficiency and renewable energy rises by \$457 million, to \$2.3 billion, with the largest increases targeting advanced manufacturing, and vehicle and building technologies. These programmes, run by the Department of Energy, are aimed at bolstering the competitiveness of industry. “Our motto is ‘Invented in America, made in America, sold worldwide,’” says energy secretary Steven Chu.

Included in the package are increases for research in solar energy, bioenergy and fossil fuels, including \$155 million for carbon capture and storage systems. But there are reductions and shifts as well: the budget for wind power remains unchanged but is allocated mainly to offshore technologies. Spending on nuclear energy continues an ongoing move toward small, modular reactors.

The National Oceanic and Atmospheric Administration (NOAA) in Washington DC receives a boost of 3%. That isn’t enough to offset both inflation and rising salaries, but

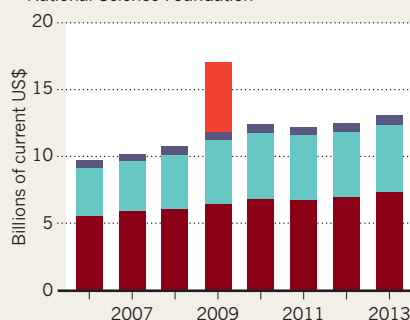
nonetheless protects a core agency priority: a programme of polar-orbiting weather and environment satellites that has been troubled by delays and cost overruns. Last year, NOAA requested a hefty increase of \$688 million to get the programme back on track, but received just under two-thirds of that. This year, the satellite programme is boosted by \$169 million.

NOAA watchers looking for signs of the

A LONG WAY TO GO

The push to double funding for three key US science agencies to boost innovation is proceeding more slowly than planned.

■ Economic stimulus funding
■ National Institute of Standards and Technology
■ Department of Energy Office of Science
■ National Science Foundation



president’s proposed reorganization of the Department of Commerce, which would move NOAA from there to the Department of the Interior, found no trace of the plan in the 2013 budget. The budget is also silent on another big initiative, the creation of a climate service within NOAA.

In what could be a third straight year of declining budgets for the US Environmental Protection Agency (EPA) in Washington DC, the agency’s funding has been slashed by 1%, to \$8.3 billion, almost \$2 billion less than in 2010. Nonetheless, funding for initiatives that target climate change and the environment rises slightly, to \$807 million, protecting core science and regulatory efforts. To compensate, the White House has cut \$359 million from a pair of clean-water grant programmes. These programmes are popular in Congress, and lawmakers have reversed similar cuts in the past.

“They did a pretty good job in making sure we are not hurting our environment and conservation programmes,” says Scott Slesinger,

legislative director at the Natural Resources Defense Council in Washington DC. But Slesinger expects Congress to inflict further cuts.

With a 3% rise for its overall budget, the US Geological Survey (USGS) in Reston, Virginia, fares better than most mission-oriented science agencies. The agency’s research and development portfolio expands from \$675.5 million to \$726.5 million. Part of the increase includes an extra \$13 million for research on the effects of hydraulic fracturing, the process used by the oil and gas industry to squeeze hydrocarbons out of non-porous rock. The president has also pumped an extra \$10.3 million into natural-hazards work, including \$2.4 million for research on quick responses to earthquakes, volcanic eruptions and landslides, and \$1.6 million to study the risk of earthquakes in the eastern United States, which was shaken by a magnitude-5.8 tremor last August.

Daniel Sarewitz, a geoscientist and co-director of the Consortium for Science, Policy and Outcomes at Arizona State University in Tempe, supports the increase for the USGS. “The survey doesn’t get a lot of attention, but it does things that are important for the nation and it’s structured in ways that make its science very useful.”

EDUCATION

The administration has taken pains to advertise a \$3-billion effort to increase and strengthen the future US science and technology workforce. For example, a combined expenditure of \$135 million by the Department of Education and the NSF aims to boost the number of science and mathematics teachers by 100,000 over the coming decade. An even more ambitious effort allocates an additional \$81 million to increasing the number of science graduates by one million — roughly 30% more than there are today — over the same period. According to Carl Wieman, associate director for science at the Office of Science and Technology Policy, simply reducing the attrition of science majors, which currently runs as high as 60%, could drive much of that increase.

Obama made a point of previewing both initiatives during the White House science fair, telling students there, “You give me confidence that America’s best days are still to come.” Now, as the budget goes to Congress, the battle to support lofty goals with real dollars begins a new round. ■



**MORE
ONLINE**

TOP STORY



India’s genetically modified cotton fiasco
go.nature.com/5mlzrn

MORE NEWS

- Did life originate in puddles on land? go.nature.com/uqpsy
- Tracking diseases through social media go.nature.com/2xasel
- Open innovation provides few malaria drug candidates go.nature.com/o4rdni

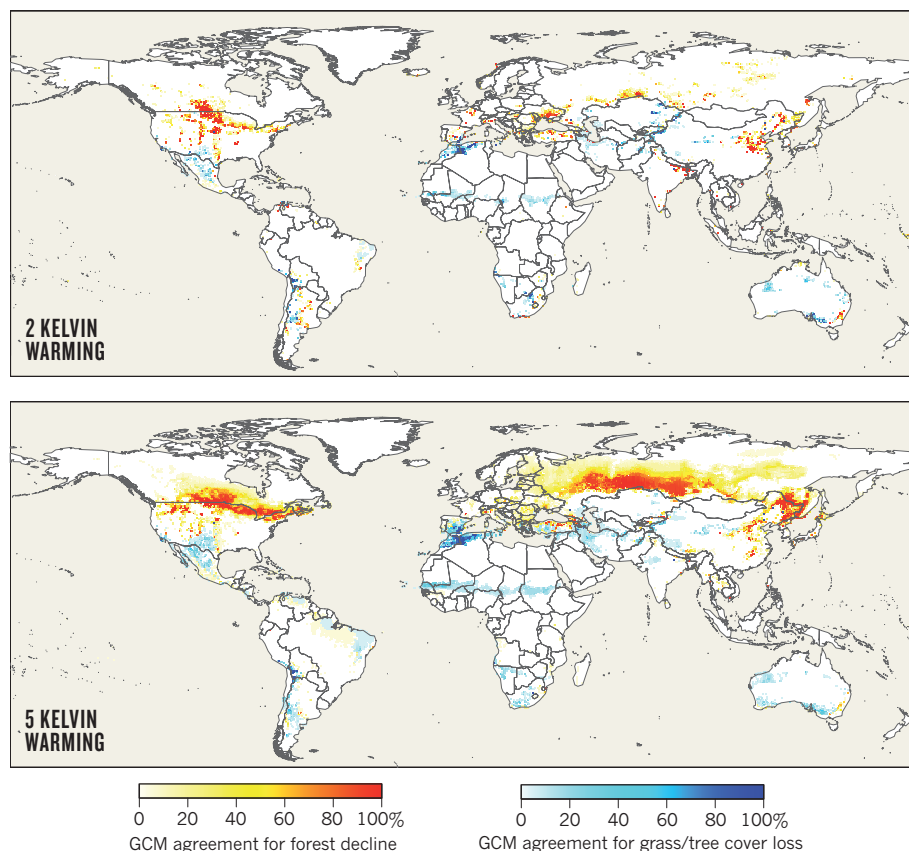
Q&A



Cristián Samper reflects on his time at the Smithsonian Institution Museum of Natural History
go.nature.com/jy1bha

THE BROWNING OF THE PLANET

A model developed at the Potsdam Institute for Climate Impact Research in Germany combined temperature and precipitation projections from 19 general circulation models (GCMs) to predict the most likely regions of vegetation loss. Results are shown for two different warming scenarios.



coordinate the fast-track programme jointly with the PIK. “Impact models have never been global, and their output is often sketchy. It is a matter of responsibility to society that we do better.”

The programme, dubbed the Inter-Sectoral Impact Model Intercomparison Project (ISI-MIP), involves more than two dozen modelling groups from eight countries, and they have set themselves a tight deadline. At the kick-off meeting at the PIK, the researchers agreed to complete a comprehensive set of model experiments within six months. All the simulations will cover the globe at the same resolution, and will be based on the same set of climate data from state-of-the-art climate models, driven by the latest greenhouse-gas emission scenarios (R. H. Moss *Nature* **463**, 747–756; 2010).

The comparison should reveal systematic biases that lead models to give widely differing results. The ‘refined’ models will still give a range of answers — but the modellers hope that the diversity will be informative rather than frustrating. “We’ll never be able to tell exactly what the future will look like,” says Ottmar Edenhofer, chief economist at the PIK. “But we can illuminate plausible paths, and multi-model comparisons help light up the black boxes.” For example, he says, the severity with which crop failures affect poor societies depends on factors such as global trade flows and local institutions and infrastructure — which some models can handle better than others.

By January 2013, the project hopes to produce papers detailing the impact of climate change on global agriculture and water supplies, vegetation and health. The results could find their way into the next report of the Intergovernmental Panel on Climate Change (IPCC), which is set to be published in 2013–14. “It will make a real difference for the assessment process,” says Chris Field, an ecologist at the Carnegie Institution for Science in Stanford, California, and co-chairman of the IPCC’s working group on impacts, adaptation and vulnerability.

The ISI-MIP will continue into 2013 and may be expanded to cover impacts on transport and energy infrastructures, both of which are vulnerable to the effects of rising temperatures and changing weather. It will also feed into other efforts, such as a five-year, 30-million-renminbi (US\$4.8-million) Chinese programme on climate-related risks, such as floods and droughts. “This exercise will greatly inform our own studies,” says QiuHong Tang, a hydrologist at the Chinese Academy of Science’s Institute of Geographic Sciences and Natural Resources Research in Beijing, and co-leader of the Chinese project. Impact modelling will be a major part of it, he says, because “climate change may greatly affect water resources and food security in the world’s most populous country”. ■

ENVIRONMENT

Models hone picture of climate impacts

International programme will improve predictions.

BY QUIRIN SCHIERMEIER

Will the warming planet be able to sustain coming generations? Few questions about the future matter more. But although modellers can forecast temperature changes and even precipitation, they struggle to say how climate change will affect the factors that make the planet habitable, such as food and water availability. Last week at the Potsdam Institute for Climate Impact Research (PIK) in Germany, researchers launched a fast-track programme to make their narratives of possible futures more coherent and useful to decision-makers.

Climate-impact models combine projections of change in physical climate with data on population, economic growth and

other socio-economic variables. For various emissions scenarios, they forecast climate-driven changes in crop yields, vegetation zones, hydrology and human health (see ‘The browning of the planet’). But they often leave out important elements: for example, models of health impact often neglect the role of social factors in spreading disease; and models of water run-off may not account for changes in water loss from plants. Researchers have built dozens of models, but have never systematically compared their performance. As a result, say critics, the literature on climate impacts is as inconclusive as it is encyclopaedic.

“Impact research is lagging behind physical climate sciences,” says Pavel Kabat, director of the International Institute for Applied Systems Analysis in Laxenburg, Austria, which is to

POLAR RESEARCH

Russians celebrate Vostok victory

Team finally drills into biggest Antarctic subglacial lake.

BY NICOLA JONES

After two decades of chilly drilling and fiery debate, a Russian team has finally broken into Lake Vostok. The largest of the lakes hidden under Antarctica's ice, and the most deeply buried, Vostok has been isolated for millions of years and may contain specially adapted microorganisms. "I'm sure they're drinking vodka this week," says John Priscu, an Antarctic researcher at Montana State University in Bozeman, who has been in contact with the Russian team.

According to Valery Lukin, director of the Russian Antarctic programme, the drill hit lake water 3,769.3 metres down at 10:25 p.m. on 5 February local time (see 'Long way down'). Temperatures were plummeting as the Antarctic summer ended, and scientists left the next day before it became too cold for planes to fly safely. "Talk about suspense. It has been a nail-biter for the past couple of weeks," says Priscu.

Although the Russian scientists have taken samples, which are most likely to be from a pocket of water just above the lake (one container was presented to Russian Prime Minister Vladimir Putin with great fanfare), they will have to wait until December to extract any frozen lake samples, and until 2013–14 to retrieve unfrozen lake water. "This is a technological achievement. The scientific pay-off is still many years away," says Mahlon Kennicutt, president of the international Scientific Committee on Antarctic Research.

The Vostok drilling project began as an ice-coring effort to examine ancient climatic conditions. By the mid-1990s, scientists had confirmed that a giant lake lurked beneath the borehole and speculated that sampling its water might yield signs of ancient life¹. By the end of the 1990s, the research community had agreed that the Vostok drilling should stop until researchers could be sure that the lake would be protected from contamination by the unsterile kerosene and Freon being used as drilling fluids. Drilling started up again in 2005 with a new plan: when the drill neared the lake, it would be replaced with a thermal probe to melt through the ice, and a plug of clean silicone fluid that would help to protect the lake water from the dirty kerosene above².

Although it is unclear whether the Russian team used the thermal probe and silicone, it

probably avoided contaminating the lake. When the drill broke through to the lake, water surged roughly 30–40 metres up the borehole, forcing 1.5 cubic metres of drilling fluid out of the top of the hole. "If everything went as they said, the only flow would be out of the lake, not into the lake," says Kennicutt.

The lake water at the bottom of the hole will freeze, and the researchers plan to drill it out next season. Previous studies found cells in samples of accretion ice³ — the bottom couple of hundred metres of the glacier made from frozen lake water — but contamination has not been ruled out. The fresh ice plug is unlikely to clear up that controversy, because the samples must be brought to the surface through drilling kerosene, says Kennicutt. The freezing process may also exclude or kill microbes, he adds.

The Russian team plans to explore the lake in 2013–14 using a variety of probes, cameras and water samplers carried down the borehole in a hermetically sealed container. One probe will measure physical conditions such as temperature and acidity, while another will carry a spectrometer to study any organic compounds in the water.

Meanwhile, the United Kingdom and the United States aim to sample water and sediments from different Antarctic subglacial lakes a year earlier, in 2012–13. Both projects will use heated glacier meltwater to bore holes that should stay open for 24 hours, a cleaner and quicker process that should allow the UK team to get through 3.1 kilometres of ice into Lake Ellsworth in just 3 days. Vostok's thicker glacier and lower temperatures would have made the process too energy-intensive to be practical there, however.

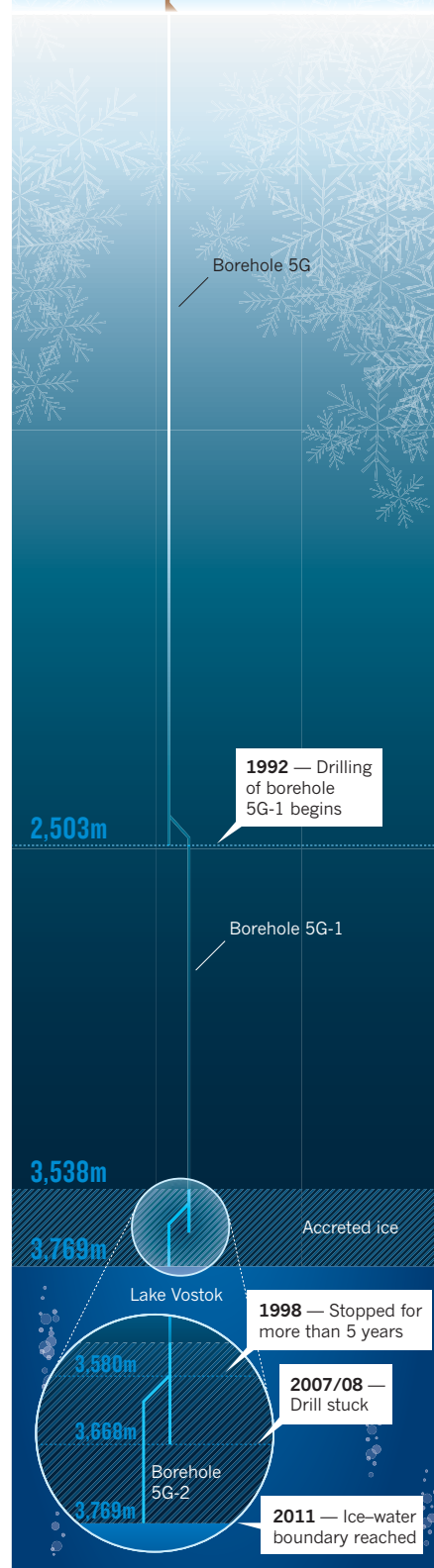
Kennicutt hopes that the Vostok, Ellsworth and US Lake Whillans projects will form the first three nodes of a network that will better sample the hundreds of subglacial Antarctic lakes. "They're not actually at the extremes of pressure and temperature, but they are limited in nutrients and energy," says Kennicutt. If life is eventually confirmed to reside in these inhospitable places, "the question is how microbes make a living down there". ■

1. Kapitsa, A. P., Ridley, J. K., Robin, G. de Q., Siegert, M. J. & Zotikov, I. A. *Nature* **381**, 684–686 (1996).
2. Lukin, V. & Bulat, S. *Geophys. Monogr. Ser.* **192**, 187–197 (2011).
3. Karl, D. M. *et al. Science* **286**, 2144–2147 (1999).

LONG WAY DOWN

About two decades after drilling at Lake Vostok began, a Russian team has finally hit water, having coped with major technical problems and contamination concerns.

1990 — Drilling begins at Vostok Station



SOURCE: REF. 2

GENOMICS

Sequencing set to alter clinical landscape

Access to whole genomes shifts potential for diagnosis, but poses challenges for doctors and regulators.

BY ERIKA CHECK HAYDEN

Sequencing a patient's complete genome can cost as little as tests that target specific genes, and in a handful of cases it has led to a life-changing diagnosis. No wonder the technology is moving fast from bench to bedside. But as the trend accelerates, researchers are left grappling with complex questions about the medical value of patient genomes, and how sequencing in the clinic should be regulated.

Two clinical sequencing programmes that launched in January, one at Baylor College of Medicine in Houston, Texas, and the other at the University of California, Los Angeles (UCLA), suggest that genomics is about to enter the clinic on a large scale. Already, a growing list of companies and hospitals offer sequencing of whole genomes or the exome (the gene-coding part of the genome). Several other efforts are set to begin this year, and many are in the spotlight this week at the Advances in Genome Biology and Technology meeting in Marco Island, Florida.

"We're all amazed by the speed at which this is happening," says Michael Watson, executive director of the American College of Medical Genetics (ACMG) in Bethesda, Maryland. "I've never seen any genetic technology move at this pace."

Sequencing sets of genes is already being used to guide cancer treatment, and many cancer centres expect to move to whole-genome or exome sequencing at some point. Clinical sequencing programmes aimed at identifying the causes of rare genetic diseases, mostly in children, are also being set up. But how often access to a patient's full genome actually leads to a useful diagnosis is an open question, because most published reports have focused on one-off success stories (see 'Put to the test').

Results from early clinical sequencing efforts are beginning to quantify these successes. Tina Hambuch, scientific liaison at Illumina, based in San Diego, California, which began offering clinical sequencing in 2009, says that a whole genome yields a diagnosis in about 40% of cases. And Wayne Grody, a medical geneticist and medical director

of UCLA's clinical sequencing programme, says that, so far, about half of the 10 patients who have come through his programme have received a diagnosis. Geneticists caution, however, that these samples are small and highly selective, and that the true rate of



PUT TO THE TEST

A growing number of successes have spurred the use of whole-genome sequencing as a diagnostic tool.

FEBRUARY 2001 Human genome draft completed by competing teams.

APRIL 2008 First sequence of an individual human, James Watson.

MARCH 2010 First sequenced family uncovers causative gene for Miller syndrome.

JUNE 2010 Doctors help to restore health of Nicholas Volker (**pictured**) after sequencing indicates that his inflammatory bowel disease could be alleviated by a bone-marrow transplant.

APRIL 2011 Sequencing spares a woman with leukaemia from undergoing a bone-marrow transplant.

JUNE 2011 Doctors report using whole-genome sequencing to improve treatment for a patient with the movement disorder dopa-responsive dystonia.

diagnosis has yet to be determined.

A further question is cost effectiveness. One speaker at the Marco Island meeting is Heidi Rehm, director of the Laboratory for Molecular Medicine at the Partners HealthCare Center for Personalized Genetic Medicine in Cambridge, Massachusetts. Rehm says that, on a standard test for known hearing-related genes, about 20% of the children who come to her clinic with deafness of unknown cause test positive for mutations in the gene that most often causes hearing loss. About the same proportion of the rest test positive for mutations in other hearing-related genes. Yet the standard tests for these sets of mutations can cost thousands of dollars. This means that, at US\$2,500–9,500, sequencing a complete genome or the exome costs about the same, and so a larger number of patients might obtain a diagnosis.

But interpreting a diagnosis is still no easy matter, because much of the variation in the genome remains poorly understood. "That's the most challenging task we face," says Rehm, who last December received one of six grants awarded by the US National Human Genome Research Institute in Bethesda, Maryland, to explore how clinical sequencing affects health outcomes in a variety of diseases.

In some instances, the diagnosis can lead to a treatment. One of the UCLA patients, for instance, was diagnosed with a rare genetic disease that is treatable by steroids. But such stories are the exception rather than the rule. "I would say a fairly rare percentage of cases are going to be like that," says Hambuch. More often, she says, sequencing reveals a rare and untreatable mutation — information that, although unwelcome, "can save the family and the medical community from doing more testing that isn't going anywhere. If you account for that, then the impact is pretty big."

Deciding what information to give doctors and patients raises its own complex set of questions, such as whether children should be told of their increased risk of an adult-onset disease, and whether doctors who do not reveal such information may later be sued for malpractice. Different clinics approach these questions in different ways. For instance, Illumina is planning to allow patients and their doctors to choose which test results they want to know, and which they do not — for example, whether or not a patient carries the genetic mutation that causes Huntington's disease, an incurable neurodegenerative disorder.

Another area of uncertainty is how regulators will enter into the equation, which could involve oversight of labs, machines, bioinformatics tools and the format of the delivered results. A number of professional groups, including the College of American Pathologists, headquartered in Northfield, Illinois, and the ACMG, are working on guidelines for the field, and some of these will be rolled out this year. ■

G. PORTER/MILWAUKEE JOURNAL SENTINEL/MCT/NEWS.COM

➔ **NATURE.COM**
Read more about
whole genomes in
the clinic at:
go.nature.com/yv7rls

EPIDEMIOLOGY

Death-rate row blurs mutant flu debate

Even if a 59% mortality rate for H5N1 is too high, the virus could still cause a flu pandemic more serious than that of 1918.

BY DECLAN BUTLER

When the US National Science Advisory Board for Biosecurity called for redaction of two papers on mutant strains of H5N1 avian influenza virus, one reason it cited was the high fatality rate of the wild virus¹. But some virologists claim that the mortality rate — 59% in officially confirmed cases (see go.nature.com/3ys4py) — has been vastly overestimated.

In an opinion article² published last month, Peter Palese and Taia Wang, virologists at Mount Sinai School of Medicine in New York, argued that the 59% figure was “driving this controversy”, but that it was “likely orders of magnitude too high”. Vincent Racaniello, a virologist at Columbia University in New York, told *Nature* that he believes any H5N1 pandemic would not come close to the 2–2.5% global mortality rate of the 1918 flu pandemic.

The virologists are reviving an old argument³ that the 59% rate, derived from officially confirmed World Health Organization (WHO) numbers of cases and deaths (see ‘Mortal danger’), does not take into account undetected cases or asymptomatic infections.

Since the current H5N1 outbreak began in 2003, researchers have done several seroprevalence surveys looking for antibodies to the virus in people from outbreak areas. Racaniello and his colleagues say that some survey results imply that H5N1 infections are common and the fatality rate is low. But all other scientists contacted by *Nature* disagree.

Malik Peiris, a flu virologist at the University of Hong Kong, says that the higher seroprevalence rates seem to be outliers. One study found that 5.6% of people tested had antibodies against H5N1 (ref. 4).

All of the others found much lower levels, and most found no detectable antibody.

“The consensus is that the rates are quite low, though not non-existent,” says Peiris.

Jeremy Farrar, director of the Oxford University Clinical Research Unit in Ho Chi Minh City, Vietnam, agrees. “Across south-east Asia we have not found much evidence of seropositivity,” he says. A review of H5N1 seroprevalence studies published last year⁵ also concluded that “transmission of the H5N1 virus from poultry to humans is rare”.

Racaniello says that the studies that found no or little seroprevalence “were simply done in the wrong populations”, and that more studies are needed.

Many scientists say that focusing on the exact numbers misses the wider picture — even if the fatality rate has been overestimated, the virus is still a severe pandemic threat. “I don’t care whether this virus has a case-fatality rate of 50% or 5% or 1%, it would still be

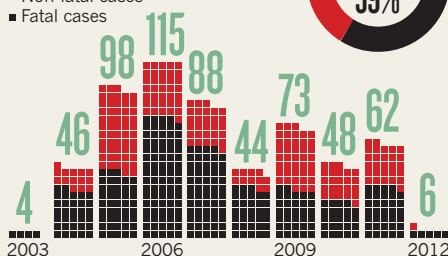
a really big problem,” says Marc Lipsitch, an epidemiologist at the Harvard School of Public Health in Boston, Massachusetts.

In fact, although the true mortality rate of H5N1 is likely to be lower than 59%, the epidemiological data suggest that it would dwarf the 0.1–0.4% rate assumed in many countries’ pandemic preparedness plans, and could far exceed that of the 1918 pandemic. Robert Webster, a flu virologist at St Jude Children’s Research Hospital in Memphis, Tennessee, warns that the H5N1 virus is a fearsome pathogen. “You walk into a poultry house that has this virus and everything is dead,” he says. “If that sort of virus were to get into humans ... my God.” ■

MORTAL DANGER

Since 2003, 584 confirmed cases of H5N1 flu have led to 345 deaths. (Numbers indicate total cases.)

■ Non-fatal cases
■ Fatal cases



➔ **NATURE.COM**

For more on mutant H5N1 flu:
nature.com/mutantflu

1. Berns, K. I. *et al. Nature* **482**, 153–154 (2012).
2. Palese, P. & Wang, T. *Proc. Natl Acad. Sci. USA* <http://dx.doi.org/10.1073/pnas.1121297109> (2012).
3. Butler, D. *Nature* **439**, 124–125 (2006).
4. Khuntirat, B. P. *Clin. Infect. Dis.* **53**, e107–e116 (2011).
5. Van Kerkhove, M. D. *et al. PLoS ONE* **6**, e14582 (2011).



CULTURAL ROOTS

A South African archaeologist digs into his own past to seek connections between climate change and human development.

BY JEFF TOLLEFSON

Metal scrapes on hard sand as archaeologist Chris Henshilwood shaves away the top layer of sediment in Blombos Cave. After just a few moments, the tip of his trowel unearths the humerus of a pint-sized tortoise that walked the Southern Cape of South Africa many millennia ago. Next come shells from local mussels and snails amid blackened soil and bits of charred wood, all remnants of an ancient feast. It was one of many enjoyed by a distinct group of early humans who visited Blombos Cave over the course of thousands of years.

The Still Bay culture was one of the most advanced Middle Stone Age groups in Africa when it emerged some 78,000 years ago in a startlingly early flourishing of the human mind. Henshilwood's excavations at Blombos Cave have revealed distinctive tools, including carefully worked stone points that probably served as knives and spear tips, and bits of rock inscribed with apparently symbolic designs. But evidence of the technology disappears abruptly in sediment about 71,000 years old, along with all proof of human habitation in southern Africa. It would be 7,000 years

Chris Henshilwood
inspects a cave on the
South African coast
near Blombos.

J. TOLLEFSON

before a new culture appeared, with a markedly different toolkit, including crescent-shaped blades probably used as arrowheads.

What drove the coming and going of these early cultures? At about the time the Still Bay culture disappeared, the globe — already in the middle of a glacial period — began to cool even further, causing sea levels to fall (see ‘Crucible of culture’). “Humans are very adaptable,” says Henshilwood, “but I think climate must have played some role in the demise of the Still Bay.”

If there is a link, it may hold broader implications. Genetic data suggest that the entire population of modern humans contracted at around the same time, then rebounded and expanded in Africa and onto other continents.

Multiple teams are now racing to determine the part climate might have played in driving human evolution during this period. Blombos Cave, with its detailed archaeological record of the Middle Stone Age, could become a key testing ground. With Francesco d’Errico, an anthropologist at the French National Centre for Scientific Research (CNRS) in Bordeaux, Henshilwood has assembled a team of archaeologists, climate modellers and palaeoclimatologists for a five-year, €2.5-million (US\$3.3-million) project to look at correlations between climate and culture during the eventful span of prehistory that includes Still Bay, and the beginnings of modern human behaviour.

“These are very daunting questions indeed, but I think they are answerable,” says Henshilwood, a native of Cape Town who now works at the University of Bergen in Norway. “If we can get some good climatic data, we can at least hazard some guesses.”

PERSONAL HISTORY

Outside the cave, a cool November breeze scours the steep slope to the shore, which Henshilwood has known since he was a child. His grandfather bought this land on the Southern Cape as a fishing retreat in 1961 and Henshilwood spent his holidays searching the hills and caves for ancient artefacts.

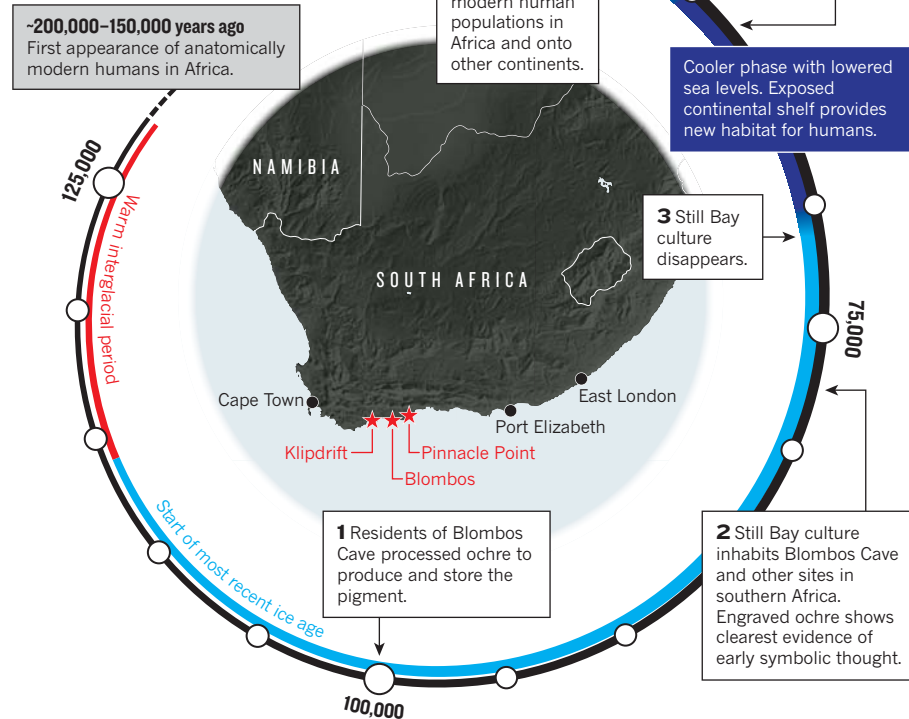
Those experiences served him well in 1985, when, out of sheer boredom in his mid-thirties, he decided to leave the family department-store business and enrol in an archaeology course at the University of Cape Town. In 1991, as a PhD student on a scholarship at the University of Cambridge, UK, he returned to Blombos in search of the same kind of artefacts that he had found as a child. What he discovered was much more significant and far older: a series of bone tools and double-sided stone points that were clearly tied to the enigmatic Still Bay period.

“It was right over there,” he says, motioning to the back of the cave. “Nobody believed us, because nobody had found a Still Bay site for 40 years.”

The Middle Stone Age was not part of his thesis, so Henshilwood covered the site up and moved on. Only in 1997 did he secure funding

CRUCIBLE OF CULTURE

During the latest ice age, human populations in southern Africa went through profound changes that sometimes coincided with major environmental shifts.



for a full excavation from the US National Science Foundation. In 2002, Henshilwood published a study¹ in *Science* documenting pieces of red, iron-rich rock called ochre, which were engraved with cross-hatched patterns. He argued that the 77,000-year-old etchings were examples of symbolic behaviour and represented the earliest known evidence of abstract thought. These and other findings have challenged the once-dominant idea that human culture — as exemplified by art such as carvings and jewellery — appeared in an explosive transformation during the Late Stone Age, some 40,000–50,000 years ago, in north Africa

and Europe. Blombos and other sites suggest a more gradual cultural and technological development, beginning far earlier, during the Middle Stone Age throughout Africa.

On a visit to Blombos in November, the cave looks like a war bunker, complete with a generator, lights and sandbags. The team has excavated just enough earth to create a workspace for a crew of five. Hundreds of steel tabs mark strata on vertical walls of sediment. While Henshilwood works on the cave's top layer, from

72,000 years ago, his partner and co-excavator, Karen van Niekerk, sifts through the bottom strata of sediments, which are roughly 100,000 years old. Centimetres away, the same layer yielded Henshilwood's most recent blockbuster find²: a toolkit of shells, grindstones and crushing stones used to process and store ochre, possibly for use as pigment or for utilitarian purposes such as tanning hides or cleaning wounds. It was further evidence that *Homo sapiens* had developed planning skills and sophistication far earlier than was once believed.

Now a postdoctoral researcher at the University of Bergen, van Niekerk has been work-

DID CLIMATE WIPE OUT THE STILL BAY CULTURE? OR DID THE PEOPLE MOVE AWAY OR PERHAPS JUST ADAPT OVER TIME?

ing with Henshilwood since the early days at Blombos. It's a good life, she says, “and a lot of work”. On this day she finishes early and heads to Henshilwood's beach house and scientific base to help a master's student, Cornelia Albrechtsen, to conduct an experiment using home-made stone and bone tools. They struggle for the better part of an hour trying to replicate the way ancient people might have opened shellfish. Then Henshilwood shows up.

“Give me one,” he says, grabbing a shell.

Within minutes, Henshilwood pops open several snails and determines which tools work best. He then departs to clean up for dinner, leaving the stunned crew to finish the experiment. "It was really impressive," Albrektson says later. "He was getting all caveman-like."

During a break in the excavations, Henshilwood stares out to sea and wonders aloud whether the Indian Ocean holds answers. Palaeoclimate records from marine sediment and ice cores suggest³ that around the time the Still Bay culture disappeared, global temperatures dropped and the polar ice sheets grew. Ocean levels fell, and the Still Bay people may have followed the sea onto the continental shelf, which would have become a productive plain.

If this idea is accurate, most of the evidence would have been submerged as the ocean returned over the past 15,000 years. Henshilwood has hiked along more than 240 kilometres of coastline in search of caves that might hold clues to the fate of the Still Bay. He hasn't found any yet, but he is beginning excavations on a site called Klipdrift Shelter, west of Blombos, that could allow him to look at the rise of Still Bay's successor: the Howiesons Poort culture, which appeared 65,000 years ago and persisted for about 5,000 years.

TIME AND TIDE

Taking a break from Blombos, Henshilwood visits the new site with Simon Armitage, a mineral-dating specialist at Royal Holloway University of London. Armitage uses a technique called optically stimulated luminescence to determine the last time a sample of dirt saw sunlight before being buried. The method requires Henshilwood and others to cover Armitage with a thick black tarpaulin and sit on its edge to prevent any light from fouling the measurements. While waiting, Henshilwood talks about the significance of the site, which has already yielded a human tooth and some artefacts with markings that could be engravings. He says the findings may turn out to be more fascinating than the decorated ochre pieces that made Blombos famous.

Once the site has been dated, the researchers will add it to environmental and cultural records from southern Africa and Europe. To construct a climate record, Henshilwood's team is sampling cave deposits, in search of clues to ancient rainfall and temperatures. They are also testing ocean sediment cores for pollen and traces of charcoal that hint at vegetation, rainfall and the frequency of fires.

The palaeoclimate data will allow a team at the CNRS to build a high-resolution model of climate in Europe and southern Africa,

beginning with the time spanning the Still Bay and Howiesons Poort cultures. The last step is to overlay the climate and cultural data onto an ecological model to analyse the environmental space occupied by specific cultures throughout time. The team can then look for links. Was one industry, for example, always associated with a particular environment? Do similar cultures occupy similar landscapes or respond to climatic shifts in similar ways?

"We can start to test our hypotheses about the role of ecology and the environment," says William Banks, who runs the modelling at the CNRS in Bordeaux.

Henshilwood and his colleagues have some friendly competition. Curtis Marean, an archaeologist at Arizona State University in Tempe, came to the cape shortly after Henshilwood, inspired by the genetic evidence of a population crash in the Middle Stone Age and thinking that the cape would have been a good place for humans to ride out hard times. He partnered with Henshilwood on a paper⁴ examining bone tools from Blombos in 2001 and went on to document the use of pigments⁵ and heat-treatment of stone tools⁶ 164,000 years ago at Pinnacle Point, less than 100 kilometres east of Blombos.

He is also looking to the sea for answers. Marean and a team of researchers have already produced an assessment³ of historical sea levels around Pinnacle Point, and now they have received money from the National Geographic Society in Washington DC and the US National Science Foundation to build a detailed geophysical map of the continental shelf. Marean thinks that the exposed shelf would have been a diverse

shrubland ecosystem with edible roots, big game for hunting and marine resources. His goal is to reconstruct the vegetation, and then use models to analyse how people might have exploited those resources.

"We need to develop a thick empirical record and put that into a really tight timescale," says Marean. "Once we have that, we can start debating the whys."

Alison Brooks, director of the Center for the Advanced Study of Hominid Paleo-

biology at the George Washington University in Washington DC, says that Henshilwood and others are producing much-needed data and hypotheses, but she warns against the dangers of oversimplification. Brooks is co-authoring a forthcoming publication that aligns palaeoclimate data with archaeological data throughout Africa, and she says that each region of the continent seems to have its own story. "There's a lot of complexity here," she says.



Stone tools from Blombos Cave.



Archaeological excavations at Blombos Cave have yielded early evidence of abstract thought.

Henshilwood acknowledges that comparing environmental and cultural data may not yield concrete answers. The disappearance of the Still Bay, he says, could have resulted from climatic change, migration, the arrival of new people or simply cultural evolution over the course of thousands of years.

Back in the cave, Henshilwood settles down into a familiar routine: digging carefully through the sediments and thinking about the past. He uncovers the remains of a clam that lives along sandy beaches and a mussel that prefers rocky shores, evidence that the Still Bay people had access to a varied coastline much like the one he has been exploring all his life. Just behind Henshilwood is another hole, carefully filled with sandbags. He dug that in 2007 as a test plot and found that the sediments inside Blombos date back at least 130,000 years, with artefacts dispersed throughout. "But that's for another day," he says, glancing at the wall of dirt in front of him. "Or another year, another decade." ■

SEE BOOKS & ARTS P.304

Jeff Tollefson covers energy and environment for *Nature* in New York.

1. Henshilwood, C. S. *et al. Science* **295**, 1278–1280 (2002).
2. Henshilwood, C. S. *et al. Science* **334**, 219–222 (2011).
3. Fisher, E. C., Bar-Matthews, M., Jerardino, A. & Marean, C. W. *Quat. Sci. Rev.* **29**, 1382–1398 (2010).
4. Henshilwood, C. S., D'errico, F., Marean, C. W., Milo, R. G. & Yates, R. J. *Hum. Evol.* **41**, 631–678 (2001).
5. Marean, C. W. *et al. Nature* **449**, 905–908 (2007).
6. Brown, K. S. *et al. Science* **325**, 859–862 (2009).

► **NATURE.COM**
For more pictures
from Blombos and
other sites, visit:
go.nature.com/2kypzy



Outside the fold

Susan Lindquist has challenged conventional thinking on how misfolded proteins drive disease and may power evolution. But she still finds that criticism stings.

BY BIJAL P. TRIVEDI

On a frigid winter's morning in 1992, Susan Lindquist, then a biologist at the University of Chicago in Illinois, trudged through the snow to the campus's intellectual-property office to share an unconventional idea for a cancer drug. A protein that she had been working on, Hsp90, guides misfolded proteins into their proper conformation. But it also applies its talents to misfolded mutant proteins in tumour cells, activating them and helping cancer to advance. Lindquist suspected that blocking Hsp90 would thwart the disease. The intellectual-property project manager she met with disagreed, calling Lindquist's idea "ridiculous" because it stemmed from experiments in yeast. His "sneering tone", she says, left an indelible mark. "It was actually one of the most insulting conversations I've had in my professional life." It led her to abandon her cancer research on Hsp90 for a decade. Today, more than a dozen drug companies are developing inhibitors of the protein as cancer treatments.

Lindquist seems able to shrug off such injustices, now. Her work over the past 20 years has consistently challenged standard thinking on evolution, inheritance and the humble yeast. She has helped to show how misfolded infectious proteins called prions can override the rules of inheritance in yeast, and how this can be used to model human disease. She has also proposed a mechanism by which organisms can unleash hidden variation and evolve by leaps and bounds. She was the first female director of the prestigious Whitehead Institute for Biomedical Research in Cambridge, Massachusetts, and has received more than a dozen awards and honours in the past five years.

In a paper being published this week in *Nature*, she and her colleagues show that in wild yeast, prions provide tangible advantages, such as survival in harsh conditions and drug resistance¹.

What is most striking about Lindquist,

however, is that despite having the self-confidence to take on controversial projects, she remains remarkably sensitive to criticism. The sting of rejection from the Chicago intellectual-property office may have dulled, but she recognizes and is dismayed by what she sees as a growing incivility among colleagues, a meanness that she thinks threatens the progress of science. "I feel like the profession is getting less and less genteel and more and more cut-throat," she says.

HEATING UP RESEARCH

Lindquist began her career at Harvard University in Cambridge, Massachusetts, in 1971, in the laboratory of Matthew Meselson, a biochemist famous for helping to show how genetic information is copied and inherited. "He was a brilliant scientist," Lindquist says, but when she started he was spending much of his time lobbying for a federal ban on biological weapons in the United States. "So he was never here."

She found the lack of a mentor very stressful in those early days. "It was terrifying and I almost left a couple of times," she says. Working more or less on her own, Lindquist decided to probe a mysterious phenomenon that researchers were exploring, called the heat-shock response. When fruitfly larvae are exposed to high temperatures, certain regions of their chromosomes 'puff up' as genes at these sites frenetically produce RNA. In work that would culminate in her PhD and eventually shape her career, Lindquist showed that applying heat to cultured fruitfly cells triggers an emergency response in which the cells manufacture heat-shock proteins, such as Hsp90, to protect themselves².

When Lindquist published her data, she says, "an awful lot of people thought it was nonsense". Colleagues dismissed the findings as an artefact — the result of heat denaturing proteins. Although the work was published in a prestigious journal, Lindquist took the criticisms hard. Her lab

Prion proteins are responsible for the colour differences in some strains of yeast.

R. KRISHNAN/S. LINDQUIST

➔ NATURE.COM
To hear an interview with Susan Lindquist:
go.nature.com/alz1ss



Susan Lindquist's career has been shaped by her investigations of proteins produced in response to 'heat shock'.

SAM OGDEN FOR NATURE

mate, collaborator and close friend at the time Steven Henikoff — now at the Fred Hutchinson Cancer Research Center in Seattle, Washington — wondered, "How can such a nice person survive" in this field?

With her newly minted PhD, Lindquist started a postdoctoral fellowship in 1976 at the University of Chicago. Two years later, the university offered her a tenure-track position. Lindquist became interested in studying heat-shock proteins in yeast, partly because it would allow her to manipulate genes more easily than in flies. A faculty member warned her against changing organisms until she had tenure, but Lindquist ignored the advice, assuming that she had little chance of getting tenure anyway. "It was really very, very difficult being a woman in science then," she says. So she pursued what she found most mysterious and fascinating.

That courage often seems to be lacking in younger scientists these days, Lindquist laments. She recalls struggling to convince students or postdocs to take on risky projects, only to learn later that when they did, their lab mates mocked them. "That shocks me," says Lindquist. She has often been afraid of being wrong — a fear that led to lots of repeated experiments — but "I didn't have a fear of a new idea".

Most of the new ideas Lindquist has developed met with resistance. In late 1993, when she proposed that a heat-shock protein called Hsp104 could untangle and dismantle clumps of protein, *Nature* initially rejected her paper. The ideas struck many as absurd, Lindquist says. "When I gave a talk about it, reactions ranged from scepticism to outright disbelief." The work was eventually published the following year³.

Still, she was literally staring at her rejected manuscript when she received a call from Yury Chernoff, then a postdoc in Susan Liebman's lab at the University of Illinois at Chicago, who had found that Hsp104 influenced a bizarre colour trait in some yeast strains. Geneticist Brian Cox, then at the University of Liverpool, UK, first described this trait⁴, called [PSI⁺], in yeast in 1965. Cox noted that when white yeast strains mate with red ones their progeny produce only white offspring, rather than the mixture of red and white predicted by conventional genetic

theory. According to one hypothesis, the trait was actually passed on not by genes but by a misfolded protein that worked like the self-replicating, disease-causing prions known to trigger fatal neurological disorders such as Creutzfeldt–Jakob disease.

Prions join together to form long, 'amyloid' fibres. Working with Chernoff, Lindquist showed how Hsp104 controls the [PSI⁺] trait by chopping up fibres of a protein called Sup35 (ref. 5). Short segments of these Sup35 fibres are passed to daughter cells and act as a template for more to form. Watching the yeast prions pass from mother cell to daughter cell was "pretty magical", Lindquist says. Moreover, the results suggested that simple yeast cells could be used to study the proteins that cause neurodegenerative disorders in humans — another idea that colleagues found hard to swallow.

For the next 15 years, Lindquist expanded her study of yeast prions. Chernoff, now editor-in-chief of the journal *Prion* and based at the Georgia Institute of Technology in Atlanta, says that Lindquist pioneered many of the biochemical and molecular techniques now

used for studying yeast prions. But her controversial hypotheses, he says, have really driven the field forward and provoked discussion and new experiments. Lindquist suggested that yeast prions are widespread and may be beneficial in some cases because they are able to switch between soluble, active states and fibrous, inactive states⁶.

Many have suggested that the prions she has been observing are artefacts of laboratory culture techniques that force proteins to behave in unnatural ways. But

in her most recent paper¹, Lindquist has shown that about one-third of the 700 or so wild yeast strains she examined harboured prions. In almost half of those strains, the prion seems to confer a beneficial trait. For example, a strain isolated from white wine is resistant to acidic environments and to the anti-fungal drug fluconazole; and a strain harvested from Lambrusco grapes is resistant to a DNA-damaging agent. When the prions in these strains are eradicated or 'cured', these useful traits disappear.

**"I FEEL LIKE THE
PROFESSION IS
GETTING LESS AND
LESS GENTEEL."**

Lindquist has also continued her studies of Hsp90. When, in the 1990s, she disabled, or knocked out, both copies of the gene that makes Hsp90 in fruitflies, the creatures died; but when she knocked out just one copy, something mysterious happened. Flies were born with a hodgepodge of physical deformities, such as shrunken or square eyes, shrivelled wings and crooked legs⁷.

Lindquist realized that Hsp90 was chaperoning proteins that contain detrimental mutations into a working form, thereby hiding their effects. Removing half the Hsp90 meant there wasn't enough of it to go around, proteins could no longer fold correctly, and the effects of all the hidden mutations became apparent. Lindquist hypothesized that the same thing happens during a natural crisis such as starvation or a change in temperature or pH. The environmental shock makes more proteins misfold and these suck up the available Hsp90, leaving a surplus of incorrectly folded proteins that could spawn the evolution of new traits. Most of this misfolding will be bad, says Lindquist. But if any of it yields a cell that is well adapted to the new conditions, some organisms could survive and thrive.

Lindquist calls Hsp90 a “capacitor” for evolutionary change. Just as an electrical capacitor stores electrical energy, Hsp90 lets hidden variation build up in the genome. When an environmental stressor trips the switch, dramatic variations can be unleashed. She found the same kinds of effects in the plant *Arabidopsis thaliana* — upturned and extra roots, exotic leaf whorls and darker hues appeared when the heat-shock protein system was put under stress⁸. Lindquist suggests that studying this phenomenon would be a powerful approach for discovering hidden variation in plants — unlocking the basis of traits such as drought resistance or salt tolerance.

Lindquist says she was unaware that these ideas would upset people. Many in the evolutionary-biology community adhere to the idea that evolution proceeds in slow, tiny steps, not the big bursts she was proposing. Nick Barton, an evolutionary geneticist at the University of Edinburgh, UK, says that the suggestion that the chaperone system releases “useful” variation when needed is controversial. “I really don't think there is much evidence for an adaptive role,” he says.

Others are more open to the hypothesis. This mechanism should be incorporated into evolutionary theory, says Massimo Pigliucci, an evolutionary biologist and philosopher at the City University of New York Graduate Center. Pigliucci says that Lindquist “put empirical meat on ideas that have been around for a while”. Still, he asks, “How important are these in the evolution of lineages?” It may take another 20 years to work that out, he says.

In August 2001, Lindquist moved from the University of Chicago to take the helm of the Whitehead Institute. It was an honour, but also a draining position that she held for only three years. She oversaw the separation of Whitehead from its genome centre, a sequencing powerhouse that had contributed much of the data for the Human Genome Project. It was a financially messy ordeal that left her desperate to focus on science, and particularly on disease-related research.

Even though she hasn't been the one to develop them, Hsp90 inhibitors have already begun to show some promise. More than 20 clinical trials are exploring their effect in cancer. “It's a hot topic,” says Len Neckers, a cancer biologist at the National Cancer Institute in Rockville, Maryland, who identified the first Hsp90 inhibitor 20 years

ago. The inhibitors might also target drug-resistant fungi that cause deadly infections in people with suppressed immune systems⁹.

Lindquist's expertise in protein folding fuelled an interest in neurodegenerative diseases. Amyloid fibres are also present in Alzheimer's, Parkinson's and Huntington's diseases, and Lindquist has championed yeast as a model for studying their effects in these conditions. In a study published last year¹⁰, she showed that a pile-up of the amyloid- β protein, a hallmark of Alzheimer's, is toxic to yeast, slowing its growth. She then used the model to screen 5,000 yeast genes for ones that might affect this toxicity. The approach was successful: it turned up 40 genes, 12 of which had human homologues, and one of which is a known risk factor for Alzheimer's. Two others interact with known risk factors.

Her hope is to pin down in yeast the initial steps that lead to amyloid formation in Alzheimer's, then to identify drugs that prevent it. The approach continues to raise eyebrows, however. “Many wondered how she could possibly model things such as Alzheimer's and Parkinson's in yeast — which are a single cell, have a short life span and, of course, don't have a brain,” says Nancy Bonini, a neurogeneticist at the University of Pennsylvania in Philadelphia.

Her grant applications have received “very mixed reviews”, says Lindquist — a charitable description, she adds. Many hardworking scientists with great ideas get their proposals turned down, she says, but she worries that the tough funding climate is dragging down the tone of grant and paper reviews. “They get exhausted, tired, and they get cranky. And then they get a paper to review.” She pauses, leans forward and says emphatically, “I think we have to stop and say, ‘No, let's not do this. Let's not be mean to somebody else because someone was mean to you.’”

Meselson, she says, instilled in her the importance of ethical and compassionate scientific conduct. It is something she has worked hard to emulate and pass on to her own trainees. In late-2010, she wrote a short commentary¹¹ entitled ‘Three quite different things that matter to me’. Think and train broadly, she wrote; be kind, be generous, don't try to destroy someone; and, have faith.

Her work and the testimonials of colleagues speak to her success with the first two, and her own words attest to the last: “When I think about my kids' future I feel very concerned,” she says, tearing up as she lists the world's environmental, social, economic and political woes. “And then I go to a lecture. I'll hear someone get up and talk about their work, and they've done something amazing. The profession that I live and breathe gives me hope.” ■

Bijal P. Trivedi is a freelance writer in Washington DC.

1. Halfmann, R. *et al. Nature* **482**, 363–368 (2012).
2. McKenzie, S. L., Henikoff, S. & Meselson, M. *Proc. Natl Acad. Sci. USA* **72**, 1117–1121 (1975).
3. Parsell, D. A. *et al. Nature* **372**, 475–478 (1994).
4. Cox, B. S. *Heredity* **20**, 505–521 (1965).
5. Chernoff, Y. O., Lindquist, S. L., Ono, B., Inge-Vechtomov, S. G. & Liebman, S. W. *Science* **268**, 880–884 (1995).
6. Alberti, S., Halfmann, R., King, O., Kapila, A. & Lindquist, S. *Cell* **137**, 146–158 (2009).
7. Rutherford, S. L. & Lindquist, S. *Nature* **396**, 336–342 (1998).
8. Queitsch, C., Sangster, T. A. & Lindquist, S. *Nature* **417**, 618–624 (2002).
9. Cowen, L. E. & Lindquist, S. *Science* **309**, 2185–2189 (2005).
10. Treusch, S. *et al. Science* **334**, 1241–1245 (2011).
11. Lindquist, S. *Mol. Biol. Cell* **21**, 3804 (2010).



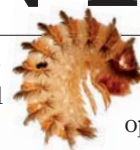
The mutations responsible for these flies' deformities are present in normal-looking flies, but their effects are usually hidden by 'chaperone' proteins.

COMMENT

PERSONAL GENOMICS Better protocols will help those with genetic disease **p.300**

PUBLIC HEALTH Call for binding global legislation to combat alcohol abuse **p.302**

ENERGY Why it can pay poor countries to keep crude oil in the ground **p.306**



EXHIBITION A museum of transgenic organisms opens in Pittsburgh **p.307**

J. HUNT, COEUR D'ALENE PRESS/AP



In many indigenous cultures, including that of the Coeur d'Alene tribe, dance is used to transmit stories and teachings on to younger generations.

Adapted to culture

Mark Pagel proposes that our ability to share and build on ideas is what made us human.

What made us human? I propose that the development of a unique capacity for culture around 200,000 years ago was the defining event in the evolution of modern humans. A fast-paced evolutionary process emerged, which by 60,000 years ago had propelled modern humans out of Africa in small tribal societies to occupy and re-shape the world in just a few tens of thousands of years.

Culture became our strategy for survival. Our ability to learn from others and to transmit and build on knowledge, technology and skills might be the most potent trait the world has seen for converting new lands and resources into more humans. Whereas other

species are confined to the environments their genes have adapted them to, we have adapted to nearly every environment on Earth.

SUMMARY

- A capacity for culture makes humans unique
- Transmitting technology and skills is our strategy for survival
- We became ultra-social through visual theft, the stealing of others' ideas
- Language evolved from a need to negotiate
- Evolution has honed the range of our talents

Humans today, I suggest, are the descendants of those who were best at using this social juggernaut to advance their interests. The defining features of our nature — our ultra-sociality and language plus various innate talents and skills — arose as adaptations to living in the prosperous social environment of human culture, not from our shared history with other animals.

Our capacity for culture rests on two building blocks that together create an unbridgeable gap in evolutionary potential between us and all other species: social learning and 'theory of mind'. Through social learning, we can copy new behaviours merely by observing others. And ►

▶ with our theory of mind we can attribute mental states to others, allowing us to guess or understand their motives. We can then choose to copy the actions, ideas or inventions that have the best outcomes.

Both characteristics may be unique to humans. What looks like social learning in other animals could be little more than socially influenced learning that makes use of behaviours already present in an animal's repertoire¹. For instance, chimpanzees manipulate things with their hands, so when one chimp uses a rock to crack open nuts, or a stick to fish for termites, another might be drawn to playing with rocks or sticks. That might by chance lead to a nut being cracked open or a termite retrieved. The reward would then reinforce the behaviour even though there was no direct imitation.

Some birds modify their behaviour when they know they are being observed by others of their species — as if 'aware' that the observer might use the knowledge it gains to its advantage. Thus, when a nutcracker bird sees another bird watching it while it hides its food, it will return alone later to hide the food in a new spot. This behaviour, also seen in other corvids, is intriguing, but it may just be a predisposition to respond to a learned behaviour; there is no good evidence beyond humans for a theory of mind^{2,3}. In fact, most human two-year-olds show a greater understanding of others' beliefs than even adult apes do.

The upshot is that although some animals seem to have what we might call cultural 'traditions' — birds pecking at milk-bottle tops to get cream, for instance, or chimpanzees cracking open nuts with rocks — these habits do not evolve or improve over time³. Even after a million years, they will still be using the same techniques — unless they acquire true social learning and a theory of mind⁴.

By comparison, human societies evolve steadily through cumulative cultural adaptation. Our knowledge, skills and technologies accumulate improvements and produce variety, as people imitate each other, choose from and modify existing forms and combine objects to make new ones⁵ — when a shaped club was combined with a hand-axe, for example, the first hafted axe was born. The result is complex and varied culture that resembles animal cultural traditions about as much as a Bach cantata resembles a gorilla beating on its chest.

VISUAL THEFT

This capacity for improvement demanded changes that are not observed anywhere else in nature. Altruism is one example. Humans cooperate with unrelated individuals and perform acts of generosity that might not be repaid. We trade and exchange things, but we also hold doors open for people, give

up seats on trains, contribute to charities and risk our lives by pulling someone from a burning building or fighting in a war. We are oddly group-focused: happy to wear silly matching shirts to sporting events, or paint our faces in the colours of our national flag, and keenly affected by the loss of our soldiers in battle. Who can forget images of Japan's fabled kamikaze warriors? You do not see this in chimpanzees.

In the rest of the animal kingdom, cooperation is generally confined to help-

"Our species's history is the progressive triumph of cooperation over conflict."

ing relatives. The theory of kin selection explains why: actions that support your relatives benefit copies of your genes. But this theory is mute in the face of the human propensity to help strangers. We should therefore consider humans as 'ultra-social', having broken free of the usual genetic constraints on altruism.

Why do we behave in these ways? I suggest that around 160,000–200,000 years ago our capacity for culture created a social crisis to which ultra-sociality was the evolutionary solution. That crisis was visual theft⁴ — the capacity to steal others' ideas.

Because we can learn simply by watching others, knowledge is available to everyone and cultures can evolve and adapt at great speed. But if I watch which lure you use to catch a fish or how you haft a hand-axe, I benefit from your ingenuity as much as you do, possibly even more, because you had to spend time tinkering before you arrived at the solution I am now copying. I might even catch that fish before you do.

Thus, once a species acquires social learning, it becomes advantageous to keep the best ideas secret, lest they be stolen. This

is illustrated today in our reluctance to share ideas — whether they be old family recipes, knowledge of fishing lures or new scientific or business plans — and also in our many patents and copyrights.

But hiding the best innovations would have brought cumulative cultural adaptation to a halt and caused our fledgling societies to collapse under the weight of suspicion and rancour. To avoid this outcome, we had to evolve the social rules and psychology that make it possible for people to exchange their ideas, knowledge and technology without undue fear of being exploited.

A great emphasis was then placed on demonstrating your own — and gauging others' — worthiness, because knowledge and technology were now held collectively by the social group, which wouldn't want to share them with cheats or competitors.

The many peculiar acts of altruism that describe our ultra-social nature evolved as costly ways for us to demonstrate our commitment, and thus our worthiness, to our cooperative group. The clearest way to show others that you are an altruist is to behave altruistically. The good reputations we earn attract altruism from others, which in turn grants us access to the material and social rewards of our communities.

We take our ultra-sociality for granted, but once such a system got going we had no choice but to become altruism 'show-offs', to compete with others for a slice of the cooperative pie. Our ultra-helpful nature is the altruism equivalent of a peacock's tail, except that the peacock uses his tail to attract a mate — we use our altruism to secure the spoils of cooperation.

Other unique features of our psychology, including our norms and morality, our expectation of fairness and our tendency towards 'moralistic aggression'



Humans are uniquely group-focused: many will deck themselves out in their team's colours.

A. YOUNG-JOON/AP

— punishing people who violate social rules — are emotions and social mechanisms we evolved to police those who might be tempted to exploit this fragile cooperative system.

SOMETHING TO TALK ABOUT

Human language differs from the grunts, chirrups, roars, odours, chest thumping and colourful displays of the rest of the animal kingdom in that it is compositional. We speak in sentences made from discrete sounds — words — that take the role of subjects, objects and verbs. Some animals make noun-like sounds — vervet monkeys can signal the approach of ground-based versus aerial predators — but only humans have been proven to use sentences.

Why? A number of ancient features of our anatomy and behaviour, such as our finely coordinated facial muscles or our primate tendencies to gesture, might have contributed to elements of our language⁶. But they do not explain why it evolved. I suggest that the complicated forms of cooperation and exchange we evolved to defuse the crisis of visual theft demanded a social technology for handling our deals, for coordinating our activities, for negotiating agreements and for broadcasting our reputations⁴. Language is that piece of social technology.

We acquired language because we were the only species with enough to talk about to pay for this expensive apparatus and the time and energy it takes to learn to use it. Lacking our social complexity, other animals don't need language, but human societies probably could not exist without it.

Even the simplest acts of exchange depend on language. Imagine you are good at making bows and I am good at making arrows, but our species has no language. I give you some arrows hoping you will give me bows in return. But you smile and, thinking my arrows are a gift, take them and walk off. I chase you, a scuffle ensues and I get stabbed with one of my own arrows. Now replay that scene with both actors having language: a cooperative and peaceful deal can be reached.

Research shows that the Neanderthals had the same version of a segment of DNA, known as *FOXP2*, that we do, and that has been implicated in the fine motor movements we use for speaking, leading many to suggest that Neanderthals, too, had language. Yet little in the archaeological record points to cumulative cultural adaptation in the Neanderthals⁷ — no musical instruments, no art, no fish-hooks or spear throwers. They did not even sew clothes. From the rule of visual theft, I suggest this dearth of culture tells us that the Neanderthals did not have language. Their human-like *FOXP2* might have given them better communication abilities than other



Our ability to build on and modify inventions has given us a selective advantage.

mammals, but in explaining the appearance of language we must look for the need for it, not just for pieces of anatomy or genes — some birds, for example, can mimic human speech, but do not share our version of *FOXP2*.

DOMESTICATED BY CULTURE

Humans have a surprisingly large range of abilities. Some of us are good at music, others at mathematics, design, language or sport, and all of these have been shown to have a significant genetic component⁸. Now, natural selection is the process by which some genetic varieties survive at the expense of others. It favours melodic singers among songbirds, and fast runners among lions and their antelope prey — poor singers remain lovelorn (and childless) and slow runners hungry or dead. We might therefore expect differences among us to get erased by natural selection. How, then, can we explain the diversity of human skills?

I believe that this variety is yet another consequence of our capacity for culture. Once our cooperative systems made it possible for people to exchange skills, goods and services, those who specialized at what they did best would have had the most to trade with others. In no other species is this possible, because no other species practises such a division of labour among unrelated individuals.

Our cultures domesticated and sorted us by our various talents, encouraging the skills to co-exist⁴. It is a scenario we should recognize, having inflicted it onto countless domesticated animals, notably

dogs. Breeds ranging from chihuahuas to Newfoundlands bear the genetic marks of having evolved specialized temperaments, skills and morphologies in response to the social environment of human whims.

Our genes might have been equally content to specialize to the opportunities our societies created, and if so, this could have implications that are relevant for contemporary society. Most of us support the societal goal of ensuring equality of opportunity. But if people have different innate skills, then such a policy could produce a 'genetic meritocracy', a society differentiated by innate predispositions.

THE MODERN WORLD

There is evidence of an upturn beginning around 40,000 years ago in the degree of positive selection acting on our genes⁹, and involving hundreds of them¹⁰. It may not be an accident that this coincided with a flourishing of human culture as seen in an explosion of artefacts, art and musical instruments, and in our occupation of the world. These fast-evolving genes constitute our wiring for culture, and they can be identified using the same methods that isolate the genes that cause medical problems.

Modern societies differ vastly from the small tribes that once competed to occupy Earth. But the old psychology plays out well in our globalized multicultural world. Our species's history is the progressive triumph of cooperation over conflict as people recognized that cooperation could return greater rewards than endless cycles of betrayal and revenge. In a diverse world, the key to promoting this cooperation is to create among people a greater sense of trust and shared values that goes beyond the highly imprecise markers of ethnic or cultural differences. This is the social glue that has fostered our ultra-sociality and can continue to do so. ■ [SEE BOOKS & ARTS P.304](#)

Mark Pagel is professor of evolutionary biology in the School of Biological Sciences, University of Reading, UK. His latest book is *Wired for Culture: Origins of the Human Social Mind* (Norton/Penguin 2012).
e-mail: m.pagel@reading.ac.uk

1. Laland, K. *Curr. Biol.* **18**, R366–R370 (2008).
2. Heyes, C. *Behav. Brain Sci.* **21**, 101–114 (1998).
3. Thornton, A. & Clutton-Brock, T. *Phil. Trans. R. Soc. B* **366**, 978–987 (2011).
4. Pagel, M. *Wired for Culture* (Norton/Penguin Press, 2012).
5. Ridley, M. *The Rational Optimist: How Prosperity Evolves* (Fourth Estate, 2010).
6. Fitch, W. T. *The Evolution of Language* (Cambridge University Press, 2010).
7. Stringer, C. *The origin of our species* (Allen Lane, 2011).
8. Pinker, S. *The Blank Slate: The Modern Denial of Human Nature* (Penguin Press, 2003).
9. Hawks, J. et al. *Proc. Natl Acad. Sci. USA* **104**, 20753–58 (2007).
10. Williamson, S. H. et al. *PLoS Genet.* **3**, e90 (2007).



Scientists are generating a wealth of human-DNA data, but no system exists for informing volunteers of their results.

Bring clinical standards to human-genetics research

Study protocols need to be rigorous, because more than science is at stake. Sometimes participants' lives depend on the results, writes **Gholson J. Lyon**.

In November 2009, I met a family in Ogden, Utah, in which three boys over two generations had died from an unknown disease with a distinct combination of symptoms, including an aged appearance, facial abnormalities and developmental delay. At the time, a fourth boy was affected; he died a few months later.

Like any researcher in human genetics and biomedicine, I wanted to identify the genes behind this disease. As a medical doctor, I heard their tragic story and drew blood from several family members in their home. Using those samples, my colleagues and I identified the genetic basis of this disease¹, which we named Ogden Syndrome after the town in which the family lives.

Then, in November 2010, another family member told me that she was four months pregnant — and she was having a boy.

She was, understandably, very worried that she might be a carrier of the mutation — two of her sisters had already lost one boy each to this heartbreaking disease. My colleagues and I had sequenced her DNA for our research, and the data suggested that she was a carrier, implying a 50% chance that her son would be

born with Ogden Syndrome. But when she asked me what I knew, I hesitated.

I was not her physician; I was a researcher, and I had done this work on a research basis, not following the specific protocol required for performing validated clinical or diagnostic tests. I couldn't be totally sure that her individual results were accurate. Should I share them with her anyway, knowing the devastation they could cause? What if I was wrong, and she terminated the pregnancy?

Now is perhaps one of the most exciting periods in human genetics and medicine — it is possible to sequence most of an entire human genome for less than the cost of many tests and procedures done routinely in clinical medicine, including magnetic resonance imaging scans and many types of surgery.

But this rapid expansion is shining a spotlight on the problems with how that information is handled and processed. Specifically, researchers are largely unable to share their findings with the people who make that research possible: study participants. At the moment,

human-genetics researchers operate in a totally unregulated environment, following their own protocols to obtain, store, track and analyse DNA — creating many opportunities for error. Researchers take shortcuts to save time and money, given that most never expect (as I did not) that their results might have a direct effect on the life of another human being. But when the result can mean the difference between life and death, mistakes are not an option.

I suggest that we change the way we collect and process samples for human-genetics research. We should create a formalized protocol akin to the rigorous process that doctors and other health-care workers go through during any clinical lab test, which practically eliminates the chances of mistakes and mix-ups. In this way, when participants want to know what we know, we will feel confident that what we tell them is correct.

In 2009, after finishing my clinical training, I moved to one of the best places in the United States for the genetic study of large pedigrees: Utah. I began to collect DNA samples from families with neuropsychiatric disorders, including individuals with severe

NATURE.COM
Read more about
Ogden Syndrome:
go.nature.com/rhitjz

developmental delay, mental retardation and autism. I also began to understand the problems with how human-genetics research is conducted.

WHEN THE UNEXPECTED OCCURS

Towards the end of my first year in Utah, I began sequencing DNA from a family in which a father and two sons were affected by severe attention-deficit and hyperactivity disorder (ADHD). Before I had finished my analysis² of the sequencing data, one of the sons revealed to me that he had a severe case of anaemia. Even though I was searching for the genetic cause of ADHD, as a physician and a human being, I felt an ethical and moral obligation to try to figure out whether he had any mutations that could have led to his anaemia. It turns out that he did.

But I was not able to return any results to him, because this research was not performed in a clinical environment. Wouldn't it help him to know that the jaundice and other problems he had battled for the past 20-plus years of his life were caused by two rare recessive mutations? Most importantly, as he moved forward in his life, wouldn't this information help him to decide with a future partner whether to undergo genetic counselling, and perhaps even genetic testing, before conceiving any children?

In the United States, all clinical laboratory testing performed on humans is regulated by the US Centers for Medicare & Medicaid Services in Baltimore, Maryland, through the Clinical Laboratory Improvement Amendments (CLIA).

When a clinician orders a blood test for anaemia, that blood is drawn by a licensed phlebotomist in an accredited laboratory setting, and the sample tube is barcoded immediately, thus reducing to about zero the chances of mix-up. The blood sample is then processed in an accredited laboratory with reagents that are carefully documented and maintained, so that haemoglobin and haematocrit are assessed and calculated in the same way for that sample as for all other samples in that laboratory, each and every time.

Even companies that perform direct-to-consumer genetic testing, such as 23andMe based in Mountain View, California, track the saliva samples quite carefully from the moment the tube is closed, so that the results can be returned to the consumer.

Now, how do most scientists in the United States conduct human-genetics research? Not in the manner described above, and not under regulation by CLIA. Instead, blood is drawn by just about anyone who is able, and there is certainly no "treating physician" ordering the blood draw (that is, someone to be held medically and legally responsible if something goes wrong or is missed).

Sometimes the sample tubes have barcodes;

sometimes they have only hand-written labels. Often, the researchers themselves extract the DNA, using standard reagents or a 'kit' available from many different companies, but there is rarely any tracking of the reagents used. DNA is sometimes extracted at a core facility using one of any number of methods, and the transferral of the samples to the core facility requires that the tubes be passed from researcher to researcher, increasing the chances of human error.

There is also extreme variability in how DNA samples are used, managed and stored. Some researchers might handle the same samples again and again, thus increasing chances for mix-ups or cross-contamination. Indeed, authors of human-genetics papers commonly eliminate samples that they suspect were mixed up. Some researchers store samples in a centralized biobank, but others use any freezer in the lab, creating many opportunities for error. There are no mandated guidelines for handling human-DNA samples in a research setting, precisely because such research is not regulated.

I never expected that a research subject would tell me that he probably had a genetic condition besides the one I was studying at the time. Some people might argue that I shouldn't have looked for the cause of the anaemia, but to me, it seems ethically and morally wrong not to try.

In the case of the Ogden family and the family with ADHD, I labelled samples by hand and gave them to a core facility at the University of Utah in Salt Lake City for DNA extraction. Although I was confident that I had performed each step as rigorously as possible, none of the reagents were tracked in any way similar to a clinical lab test. As such, the results did not meet the very high standards required of clinical tests.

I have since asked the physician of the man with ADHD to follow up my results with a CLIA-certified genetic test to confirm that the man carries the anaemia mutations, so that this clinician can release the information to him. Even now, many months later, the testing has not been performed, because most clinicians face roadblocks such as finding an available gene diagnostic test, dealing with insurance and arranging for appropriate genetic counselling.

In the case of the woman who I suspected was carrying a male fetus with Ogden Syndrome, I faced a major dilemma owing to time constraints. Given that my results could have been incorrect, and might have caused undue stress and possibly even an unnecessary termination of the pregnancy,

I chose to "first, do no harm" — I did not return my research result to her, and I instead attempted to validate it in a CLIA-certified lab at a major diagnostic facility. It was a long and bureaucratic process, but after several months, in July 2011, we had a formal, CLIA-certified genetic test for the specific mutation in *NAA10*, the gene associated with Ogden Syndrome.

Unfortunately, by that time, the woman had given birth to her son. As I had feared, he had the disease. Sadly, he died in June 2011, four days before the paper describing the mutation that killed him was published.

EXPECT THE UNEXPECTED

There are increasingly limited resources for biomedical research, and it can take 20 years or more to translate genetic discoveries into new drugs or other treatments. So why not help the families and research participants now, by deriving the highest possible value from every DNA sample we sequence?

Participants want to be involved in the research process and be told about any medically important findings. I am therefore suggesting that the entire process of DNA collection and genome sequencing for humans could and should be performed in a proper clinical environment, so that physicians can immediately return all relevant genomic information much more easily, and perhaps even link such information to medical records so that it is available for re-analysis as our knowledge expands.

This means establishing suitable guidelines for the collection, tracking and sequencing of DNA samples from human participants, along with training health-care professionals in genetics counselling.

To make these changes possible, grant agencies should consider setting aside funding to establish clinically certified protocols for handling human genomic data, including findings perhaps unrelated to the original research goals³. After all, these agencies are supported by taxpayers, and the data ought to be given back to those donating their time and DNA to the research. We cannot forget the wise words of the late geneticist Charles Epstein, from his 2001 William Allan award lecture: "the operative word in 'human genetics' is 'human.' Human genetics is about human beings — about humanity and humaneness."³ ■

Gholson J. Lyon is in the Institute for Genomic Medicine, Utah Foundation for Biomedical Research, Salt Lake City, Utah 84106, USA.
e-mail: gholsonjlyon@gmail.com

1. Rope, A. F. *et al.* *Am. J. Hum. Genet.* **89**, 28–43 (2011).
2. Lyon, G. J. *et al.* *Discov. Med.* **12**, 41–55 (2011).
3. Epstein, C. J. *Am. J. Hum. Genet.* **70**, 300–313 (2002).

Regulate alcohol for global health

The World Health Organization is the only body that can promote health through the use of international law. It should make alcohol its next target, says **Devi Sridhar**.

Unlike any other global-health body, the World Health Organization (WHO) can create legally binding conventions, and it only requires a two-thirds majority vote to do so. Yet this power is vastly underused. In more than 60 years, this United Nations agency has produced only two major treaties: the International Health Regulations, which require countries to report certain disease outbreaks and public-health events; and the Framework Convention on Tobacco Control, which commits governments to making legislative moves to reduce the demand for, and supply of, tobacco. The WHO has shown a reluctance to use hard legal instruments. Instead, it tries to influence societal norms through guidelines and recommendations¹. This is a major missed opportunity.

Now is the time for the WHO to take a bold step and move towards a third treaty to protect world health. There is an obvious target. About 2.5 million deaths a year, almost 4% of all deaths worldwide, are attributed to alcohol — more than the number of deaths caused by HIV/AIDS, tuberculosis or malaria². Alcohol consumption is the world's third-largest risk factor for health burden; in middle-income countries, which constitute almost half of the world's population, it is the greatest risk (see 'Health burdens').

There are some good, evidence-based efforts for alcohol control already in place, such as the 2010 WHO Global Strategy to Reduce Harmful Use of Alcohol. This document lays out ten areas in which action can be taken, from raising awareness to preventing drink-driving and restricting the availability, marketing and pricing of alcohol. Its recommended policy interventions are general and sensible, including: banning unlimited drinks specials; enforcing a reasonable minimum age limit for purchase; and enacting graduated licensing for novice drivers with zero-tolerance for drink-driving. The strategy helpfully summarizes the cost-effectiveness of various strategies. But this is a portfolio of useful information and policy tips, not a binding document.

A WHO Framework Convention on Alcohol Control could and should turn those recommendations into legal requirements for member states. What difference

would this make? Formally, countries would commit to applying the agreement through national legislation — which would require a ream of new policies for nations such as India where current regulation isn't so comprehensive. Nations would be required to report to the WHO on their progress. The international community would have a shared responsibility to support these efforts by providing financial and technical assistance as needed. Informally, ministries of health would have a stronger domestic negotiating position in prioritizing alcohol regulation above economic concerns.

domestic public health. Despite a binding Framework Convention on Tobacco Control, tobacco use is increasing in many poor countries, and is still the second-largest cause of disease risk in middle-income countries. The problem is that oversight is minimal and no strong enforcement mechanisms exist, so compliance is weak.

STRENGTHENED POWERS

To help overcome such problems, the WHO should endorse a commission on global health law, headed by an independent expert. Through analysis of other regimes,

such as those of trade and finance, that have arguably been more successful in utilizing international law, this commission could provide recommendations on how to strengthen the WHO's normative power.

The WHO's legal potential should not be focused solely on individual health hazards such as alcohol and tobacco, it should be used to create a broad framework convention on global health³. This would identify a basic package of health services that governments ought to provide; identify who would be obliged to provide what; and examine how this could be achieved through reform of global health governance.

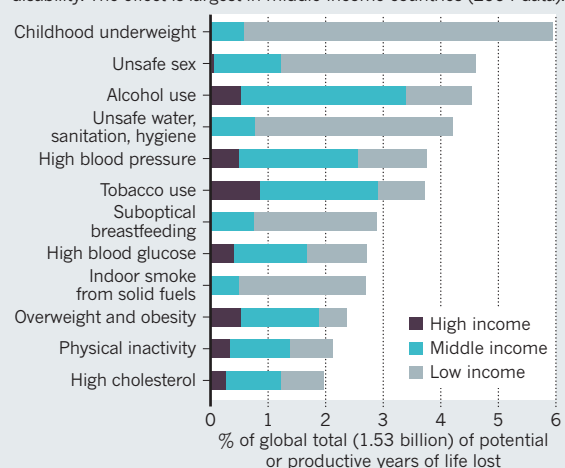
To flourish in an environment with numerous other better-financed and more-inclusive institutions, the WHO must take a hard look at itself and what makes it special. Other

bodies can provide technical advice, give money, influence domestic health policy, assist in development and advocate for the importance of health in government policy. The WHO is the only body with the legitimacy and authority to proactively promote health through the use of international law. It needs to do so. ■

Devi Sridhar is in the Department of Public Health and the George Centre for Healthcare Innovation, Wolfson College, University of Oxford, Oxford OX2 6UD, UK.
e-mail: devi.sridhar@dph.ox.ac.uk

HEALTH BURDENS

Alcohol is the third-largest risk factor for loss of years to disease and disability. The effect is largest in middle-income countries (2004 data).



Non-governmental organizations would be able to pressure governments, and even bring issues to court.

The creation of a framework convention requires much political work and preparation. The WHO secretariat should, for example, map out the positions of countries on alcohol use, their links to industry, and how best to overcome opposition in each nation. Doing so will require donor funding for a special cabinet project, as was done for tobacco. The overarching goal would be to assemble a 'coalition of the willing and able' to push this agenda forward in the World Health Assembly — the WHO's decision-making body.

We should not be overly idealistic about the effect of international health law on

► NATURE.COM
The WHO needs to be reformed:
go.nature.com/snx4vp

1. L'hirondel, A. & Yach, D. *World Health Stat. Q.* **51**, 79–87 (1998).
2. World Health Organization. *Global Status Report on Alcohol and Health 2011* (WHO, 2011); available at <http://go.nature.com/ymav6z>
3. Gostin, L. O. et al. *PLoS Med* **8**, e1001031 (2011).



With the advent of ancient cave art, human cultural innovation burst into being.

EVOLUTION

Custom built

Culture is both a product and a driver of human evolution, finds **Peter Richerson**.

When Charles Darwin discussed the forces driving human evolution in his 1871 book *The Descent of Man*, he placed cultural change — mostly under such terms as traditions, customs and inherited habits — on an equal footing with organic evolution. Darwin's idea that cultural traits adapt, change and experience selection as they are passed within and between generations attracted important followers among early behavioural scientists, but he had almost no influence on the anthropologists, sociologists and historians who dominated the study of human culture for most of the twentieth century.

Yet the parallels between culture and biology had long been obvious — just look at the tree diagrams for language relationships and for related species. Evolutionary biologist Mark Pagel's pioneering contribution has been to show that this similarity is more than skin-deep, and that methods for revealing the evolutionary history of genes can illuminate the historical relationships between languages and other culturally transmitted behaviours.

Wired for Culture provides a wide-ranging, non-technical survey of the field. Pagel's two main themes are the role of cultural evolution in patterns of cooperation and its part



Wired for Culture:
Origins of the
Human Social
Mind/The Natural
History of Human
Cooperation

MARK PAGEL

W. W. Norton/Allen
Lane: 2012. 432 pp.
\$29.95/£25

in human economic diversity — our ability to make a living in various ways in all kinds of environments. Economic diversity is a major reason for our success as a species. Modern humans first used culturally transmitted toolkits to underpin an expansion out of Africa around 60,000 years ago, in the late Pleistocene epoch. As humans dispersed, populations underwent a series of cultural

adaptive radiations as they found differing ways to exploit their local wild resources and to domesticate plants and animals.

These changes required parallel innovations in social organization. In hunter-gatherer economies, between a few hundred and a thousand individuals cooperate to finesse risk and sustain a modest division of labour. Agricultural and industrial societies cooperate on far larger scales than hunter-gatherers to sustain an intricate system of exchanges between specialized workers with complementary skills. These workers are often unknown to each other, yet are utterly interdependent.

Art and religion mobilize our emotions in support of collective projects. In Pagel's terms, art and religion can act as enhancers that promote adaptive behaviour, but in other times and places, they can be selfish mind drugs — the cultural analogues of microbial pathogens. We use our extraordinary linguistic systems to operate these social structures by articulating morals through myths and stories, debating constitutions, negotiating deals, making requests, giving orders and making or breaking reputations through gossip.

Pagel's story of culture's dominant role in human evolution is vivid and effective. Inevitably, he simplifies some concepts in order to craft an accessible, coherent narrative. For example, he portrays Neanderthals as the hapless victims of invading modern humans who used more sophisticated technology. In fact, much evidence suggests that modern humans usually made the same sorts of tools as Neanderthals, and the question of why this species disappeared is still quite open.

Pagel has an interesting take on what is perhaps the deepest controversy among Darwinists: does culture fundamentally change the evolutionary dynamics of our

species? As biologist Kevin Laland and his colleagues wrote last year in *Science*, many evolutionists take

NATURE.COM
For more on
cooperation:
go.nature.com/rnqxac

culture to be a proximate phenomenon — a mechanism for achieving something, rather than the ultimate, evolutionary reason for that achievement.

The classic example of this distinction between 'how' and 'why' in biology is bird migration. The proximate causes of migration are hormonal responses to changing day length that motivate and prepare birds for long-distance flight. The ultimate cause is selection for migratory behaviour, which in turn results in genetic change. A common inference is that causation in biology flows one way — from selection, to genetic change, to traits adapted to their environment. But human culture raises two issues regarding the arrow of causation.

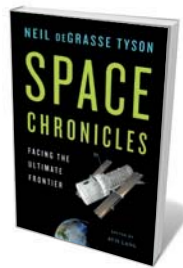
First, there are reasons to think that genetic change might be a response to cultural shifts as well as their cause. Pagel certainly thinks so. For example, he speculates that division of labour in humans has created selection for innate differences in personality and talent. Even the simplest human societies divide work between men and women and then pool the results of their specialized labours. In complex societies, myriad occupations trade their specialized products to assemble the goods that each person needs. Pagel also reviews evidence that many genes came under selection after we spread out of Africa and diversified our economic systems.

Second, culture creates an arena for selection and inheritance that is separate from genes. For example, as Pagel emphasizes, much cultural variation exists at the tribal, rather than individual, level. Some of us, including myself, think that selection of cultural differences between proto-tribes at the group level — with successful groups displacing their neighbours' cultures through the spread of people or their ideas — might have led to selection of genes favouring docility, empathy and group loyalty.

Pagel briefly discusses this concept, but favours the view that even suicidal self-sacrifice can evolve as a result of self-interested benefits to individuals, rather than to their groups. He puts great weight on systems that use reputation to monitor and enforce good behaviour, arguing that the benefits of a good reputation, and the effects of a bad one, more than outweigh the personal costs of helping other members of your group. The resolution of this issue is perhaps the most important task in the study of human evolution, and it is a pity that *Wired for Culture* does not convey what is at stake. But scholarly quibbles aside, this is the best popular science book on culture so far. ■ [SEE COMMENT P.297](#)

Peter Richerson is a professor of ecology and human evolution at the University of California, Davis, USA. He is the co-author, with Robert Boyd, of *Not By Genes Alone*. e-mail: pjricherson@ucdavis.edu

Books in brief



Space Chronicles: Facing the Ultimate Frontier

Neil deGrasse Tyson W. W. NORTON 368 pp. \$26.95 (2012)

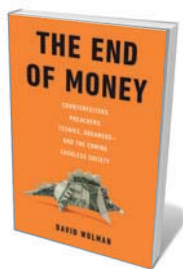
Astrophysicist Neil deGrasse Tyson is on a space mission of his own. With NASA's research and exploration now diminished, Tyson — director of the American Museum of Natural History's Hayden Planetarium — is keenly focused on what the United States will lose by failing to reinvent its space programme. In this collection, mined from 15 years of commentary and interviews edited by Avis Lang, he spotlights issues that underline the central importance of curiosity about the great beyond — from the nature of discovery to propulsion in deep space.



The Evolved Apprentice: How Evolution Made Humans Unique

Kim Sterelny MIT PRESS 264 pp. £24.95 (2012)

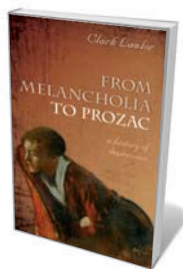
We are inescapably different from the other great apes — sexually, morphologically and socially. In a book that forms part of the Jean Nicod Lecture Series, philosopher of biology Kim Sterelny tries to answer the vexed questions of why that is by arguing that our divergence from our closest cousins over the past 3 million years is down to a gradually enriched learning environment. Cooperative foraging, he posits, paved the way for positive-feedback loops that, incrementally and over vast reaches of time, led to adapted minds that were nurtured within adapted environments.



The End of Money: Counterfeiters, Preachers, Techies, Dreamers — and the Coming Cashless Society

David Wolman DA CAPO PRESS 240 pp. £16.99 (2012)

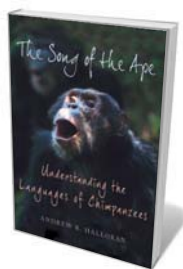
'Filthy lucre', David Wolman shows us, is a particularly apt phrase. Minting technologies gobble huge amounts of metals, cotton and water; the transport and production of cash has a giant carbon footprint; and *Staphylococcus* bacteria have been detected on 94% of US one-dollar bills. In this fascinating book, *Wired* contributing editor Wolman argues that its end is nigh. He spent a cash-free year researching everything from counterfeiting to tax-dodging and failed currencies, and looks at digital solutions such as smart banknotes.



From Melancholia to Prozac: A History of Depression

Clark Lawlor OXFORD UNIVERSITY PRESS 288 pp. £14.99 (2012)

Seventeenth-century scholar Robert Burton may have anatomized melancholy, but the definitions, diagnoses and treatments of and for depression are still hotly debated by the pharmaceutical industry, psychiatry, psychology and affected citizens. Writer Clark Lawlor trawls history from the classical era onwards, shining some light on this psychological dark horse. Along the way, he touches on radical differences in cultural definitions, explores tensions between the biomedical model and humanistic concepts, and weighs up 'cures' from talking therapy to drugs.



The Song of the Ape: Understanding the Languages of Chimpanzees

Andrew R. Halloran ST MARTIN'S PRESS 288 pp. \$25.99 (2012)

A chance observation drove primatologist Andrew Halloran to study how chimpanzees communicate. While keeper to a group of island-bound chimps, he saw five ousted members calmly using a rowing boat to escape, as if they had discussed a plan. Examining the histories of these five, he picked out and recorded dozens of phrases in their vocalizations. Weaving in the stories of controversial attempts to teach sign language to primates such as Nim Chimsky, Halloran concludes that chimps have their own highly complex lexicon.

ENERGY

A crude awakening

John Vidal is gripped by a book that reveals how natural riches can impoverish nations.

When oil was first extracted and sent from Lago Agrio, Ecuador, in July 1972, the military dictatorship held a ceremony signalling a new era. People dipped hands in the yellow-brown crude, children were baptized with it, and the first barrel was placed in a museum in Quito.

Forty years later, Ecuador has extracted nearly half of its known reserves, earning it US\$130 billion. That has paid for needed infrastructure, but has also led to widespread corruption, impoverishment, inequality, insurgency and environmental devastation. Crude oil has transformed Ecuador — just not in the way Ecuadoreans expected. The situation could be shifting there, but is a familiar one in the developing world. Political scientist Michael Ross devotes *The Oil Curse* to unpicking it.

This 'paradox of plenty' afflicts as many as 40 other developing countries, among them Nigeria, Cameroon and Equatorial Guinea. So far, only a handful of countries have avoided it, including Norway, Britain and a few smaller Arab states. Analysts have struggled to explain the paradox ever since British economist Richard Auty recognized it in 1993, invoking reasons ranging from foreign oil companies creaming off big profits to a simple lack of readiness for sudden riches.

Ross largely dismisses such triggers, suggesting that the fault lies mostly in the nature of oil wealth itself. Modern oil revenues, he proposes, have a more powerful and harmful effect on poor countries than money from other minerals because the sums involved are huge (now consistently more than \$100 a barrel), do not come from taxing citizens and are easy to conceal from public scrutiny.

Poor governance, Ross says, does play a part: oil-funded rulers can use 'petro-dollars' to block democratic reforms, an argument backed by stories from the pro-democracy uprisings of the Arab Spring. Protesters in oil-poorer countries such as Tunisia and Egypt found it easier to overthrow their rulers than did those in oil-rich states like Saudi Arabia, Libya and Algeria, he notes.

Ross is less convincing in tracing the start of the 'oil curse' back to the early 1970s, when prices quadrupled in a few months and many governments seized control of their countries' oil industries. Before nationalization, he argues, oil-rich developing countries were not that different from others. His research



The Oil Curse:
How Petroleum
Wealth Shapes the
Development of
Nations

MICHAEL L. ROSS
Princeton University
Press: 2012. 296 pp.
\$29.95, £19.95

shows how such countries are today more likely to be ruled by autocrats and to descend into civil war than countries with no oil reserves. Oil wealth also creates less economic growth than it should, and produces more work for men than women.

These are all good points, but Ross's exoneration of corporate and colonial powers before the 1970s weakens his argument. To many developing countries, an oil curse is just an escalation of colonial pillage. Oil, along with land acquisitions, is simply the latest resource to be taken by rapacious companies and national elites, leaving the majority of citizens as bystanders in the development process.

Ross bravely suggests remedies, but I fear that most will be dismissed as impractical or naively neo-liberal. Because he attributes the malaise partly to state ownership of oil assets, he advocates some level of privatization of oil industries. He skips over President Hugo Chavez's 2007 nationalization of Venezuela's oil reserves to pay for social reforms, and fails to ask why any poor country would be allowed by its people to sell off its major asset.

In searching for solutions, Ross looks at reducing the size of petroleum revenues and imposing sanctions against "undemocratic" governments — but without taking into account the anger and global price rise this could provoke. He also urges governments and corporations worldwide to be more transparent, although critics might doubt that all would take him up on it.

Ross wisely advises countries to distribute oil wealth directly to citizens. And he is useful on 'barter' contracts between nations, which avoid corruption by exchanging oil for goods or resources rather than money. But, sadly, he touches only briefly on the idea of leaving the oil in the ground, and doesn't mention Ecuador's radical experiment along those lines, initiated by President Rafael Correa in 2007.

Correa has asked the world to pay Ecuador not to extract oil reserves worth around \$7.2 billion from a 1,200-square-kilometre block of the Yasuni national park. Some 20% of Ecuador's remaining reserves are to be left untapped in return for around half of the revenue they would have been worth if exploited. In late 2011, with the full backing of the United Nations Development Programme and the polled approval of the Ecuadorean people, Correa pledged to "leave the oil in the soil". The first \$100-million tranche, to be used for conservation and renewable-energy projects, has been lodged with the UN.

Economists have mostly shied away from full costings of the ecological and social devastation of oil use. Were they to do so with the thoroughness and authority displayed by Ross in *The Oil Curse*, they might start to develop the new economic model for oil and other extractive industries that is so desperately needed. ■

John Vidal is environment editor of The Guardian newspaper in London.
e-mail: john.vidal@guardian.co.uk



Revenues from oil fields in developing nations such as Nigeria are rarely passed directly on to citizens.



A dermestid beetle that dined on transgenic mosquitoes earned a place in Richard Pell's museum.

Q&A Richard Pell

Transgene curator

Next month in Pittsburgh, Pennsylvania, artist Richard Pell opens the Center for PostNatural History — a museum of bioengineered organisms. He talks about the joys and pitfalls involved in collecting genetically modified maize, mosquitoes and zebrafish.

NICK HIGGINS

Why did you start the museum?

As an artist, I made robots in an attempt to start an ethical conversation in the engineering community about funding and other political issues. Then I was introduced to synthetic biology by one of the field's pioneers, Chris Voigt, who is now at the Massachusetts Institute of Technology in Cambridge. I began to wonder why transgenic organisms weren't shown on the evolutionary tree. So I began collecting specimens of living things that had been intentionally and heritably altered by humans.



What is the museum's focus?

The museum is essentially anthropocentric — it looks at the organisms that we alter, but also at how they alter us. The history within an engineered organism is vast, and represents the continuum of human manipulation of plants and animals. For example, the rats we breed to develop human-like tumours will shape the progress of medical research, which in turn will have an effect on which of us survive.

What specimens have you collected?

Examples include GloFish, which are zebrafish that contain genes from bioluminescent jellyfish and coral — the only transgenic organism you can buy in a pet shop. We couldn't acquire genetically modified maize [corn] directly without entering into an elaborate licensing agreement with its developers, Monsanto. But by planting maize kernels from shop-bought animal food and testing to see whether the plants survived the pesticide Roundup, we were able to add Roundup-Ready maize to our collection. We're not trying to be provocative, we're just documenting thoroughly.

What specimens don't you have?

The 'biosteel goat' designed by Canadian company Nexia Biotechnologies. It produces milk containing spider silk that could be used to make stronger bulletproof vests — one of the first 'biofactories'. The US Defense Advanced Research Projects Agency [DARPA] moved half of the herd to a decommissioned airforce base in upstate New York. The other half went to an ongoing research project at the University of Wyoming in Laramie. The status of the DARPA goats remains unknown. Our exhibit consists of a diorama of the military goat farm, based on images from Google Earth.

Does the collection include any insects?

I had a nice stock of mosquitoes, which had been altered so that they could not harbour the parasite that causes malaria. This is one of a handful of organisms engineered to be released into the wild. Last year, when I opened up my mosquito box, it was empty apart from a dermestid beetle larva, a common pest in museums. Despite my attempt to make a habitat, it died, but because its diet consisted exclusively of genetically engineered mosquitoes, it earned a place in the collection.

Why do you have an exhibit on 'genetic copy prevention'?

Companies wanting to sell living things perceive a fundamental problem: their products reproduce for free. One solution is the terminator gene patented by Monsanto, an on-off switch that allows the organism to reproduce in the lab but that makes it go sterile in the wild. There are other approaches to limiting reproduction: in the late 1950s, the United States built factories to irradiate millions of screw worms, which feed on the living flesh of livestock. Sterilized male screw worms were then dropped from aeroplanes so that wild females would mate with them. This eradicated the insect from US livestock.

Does the museum cover genetically altered humans?

We don't have exhibits on that topic yet, but we do archive this type of research. For example, in 2007, researchers at Cornell University in New York produced a transgenic human embryo that expressed a red fluorescent protein from coral, and allowed it to grow for five days before terminating it. And in gene therapy, a hacked retrovirus inserts foreign DNA into a patient's genome to produce a certain protein. That change is not supposed to be heritable, but there is the possibility that the virus could make its way into a germ cell. Although both of these examples are minor changes in comparison with how humans have altered other organisms, I think that this will be an area of interest for the museum in the future.

How does the museum deal with people's biased views?

The rhetoric around altered organisms has become narrow, both for those who are afraid of ' Frankenfoods' and those who believe that genetic engineering will cure cancer. People often want to have their own belief system mirrored in your rhetoric, or at least they want someone else's bias so that they can recognize and argue with it. Otherwise they must argue with themselves, which is uncomfortable but exactly the experience we want them to have. ■

INTERVIEW BY JASCHA HOFFMAN

Correspondence

No catch to UK charity funding

Universities receiving upwards of £1 billion (US\$1.6 billion) in total annually from UK medical research charities would probably disagree that their benefactors are 'hijacking' university resources (*Nature* **481**, 260; 2012).

The UK government supports charity-funded research as part of its higher-education funding. This enables charitable funds, often donated by the public, to be spent directly on research while the government pays universities to cover the costs of overheads and infrastructure.

Charities themselves invest in research infrastructure and resources. For example, the Wellcome Trust and Cancer Research UK are collaborating with the Medical Research Council and three London universities to invest £650 million in the Francis Crick Institute, a world-class research centre. The British Heart Foundation has committed £10 million to new university buildings since 2010, including £1 million towards the Centre for Regenerative Medicine at the University of Edinburgh.

The UK research base benefits from the breadth and diversity provided by a mix of public, for-profit and charitable funders.

Mark Walport Wellcome Trust, London, UK.

m.walport@wellcome.ac.uk

Iain Foulkes Cancer Research UK, London, UK.

Peter Weissberg British Heart Foundation, London, UK.

Delyth Morgan Breast Cancer Campaign, London, UK.

Sharmila Nebhrajani Association of Medical Research Charities, London, UK.

Whaling: don't trade the moratorium away

In their proposal to allocate 'whale shares' to both whalers and conservationists as an alternative to the International

Whaling Commission (IWC) moratorium on commercial whaling, Christopher Costello and colleagues overlook several factors (*Nature* **481**, 139–140; 2012).

Commercial whaling is in decline. In Japan, it is becoming less economically viable as consumer demand and whale-meat sales revenues fall — even with an increasing government subsidy, which this year is roughly ¥2.3 billion (US\$30 million). Demand is also waning in Iceland and Norway. A ban on international trade prevents these countries from securing new markets. Last year, the global stockpile of unwanted whale meat reached more than 7,000 tonnes.

The effective management of commercial whaling would cost a lot more than its protagonists can afford and than non-whaling nations are willing to pay.

Costello *et al.* also overlook the high costs of the independent surveys and analysis that would be needed to generate safe quotas for whaling, as well as the international compliance scheme required to enforce regulations.

The IWC's founding treaty does not allow for quotas to be allocated to individual countries. Its renegotiation to facilitate a scheme such as Costello and colleagues describe would require unanimity, which is currently unthinkable. Given all this, it would be foolhardy to trade away the moratorium.

Mark Peter Simmonds Whale and Dolphin Conservation Society, Chippenham, UK.

mark.simmonds@wdcs.org

Sue Fisher Animal Welfare Institute, Washington DC, USA.

Whaling: ways to agree on quotas

The sticking point in discussions of whale conservation schemes (*Nature* **481**, 139–140; 2012) has been reaching agreement on the total catch that each whale population can sustain. The International Whaling Commission's scientific

committee has developed and simulation-tested an adaptive algorithm, the Revised Management Procedure (RMP), to determine safe catch limits. For most populations, RMP limits are below the numbers discussed during the commission's failed 2010 negotiation of catch levels.

The whale-related expenditure of most of the organizations mentioned by Costello and colleagues goes largely to scientific research and outreach, not to protests. These funds have helped to develop tools such as the RMP, genetic techniques to monitor markets and improved methods for estimating whale numbers and demography.

If conservation organizations were to 'buy' whales, it would not necessarily reduce the numbers killed. Most catch quotas set by governments lie well above the numbers actually taken. Even if the more conservative RMP catch limits were applied to a new whale market, 'buying' a given number would not save that number over time. Under an adaptive-feedback management procedure such as the RMP, killing fewer whales one year tends to increase catch limits in subsequent years.

Justin G. Cooke Centre for Ecosystem Management Studies, Emmendingen, Germany.
jgc@cems.de

Russell Leaper University of Aberdeen, UK.

Vassili Papastavrou University of Bristol, UK.

Big data deserve a bigger audience

The huge repositories of data collected by services such as Twitter, Facebook and Google can cause serious problems beyond quality control (*Nature* **481**, 25; 2012).

Many of the emerging 'big data' come from private sources that are inaccessible to other researchers. The data source may be hidden, compounding problems of verification, as well as concerns about the

generality of the results.

These results are meaningful only if many other data sets reveal the same behaviour. This uncovers a deeper problem: if an independent set of data fails to validate results derived from privately owned data, how do we know whether it is because those data are not universal or because the authors made a mistake?

If this trend continues, we could see a small group of scientists with access to private data repositories enjoying an unfair amount of attention at the expense of equally talented researchers without these 'connections'.

Bernardo A. Huberman HP Labs, Palo Alto, California, USA.

bernardo.huberman@hp.com

Data audits could curb misconduct

Universities and government research institutes could perhaps learn from the private sector when it comes to curbing research misconduct (*Nature* **481**, 237–238; 2012).

Research entities should undergo independent audits of scientific data annually by certified public scientists, in much the same way as businesses and not-for-profit organizations are independently audited by certified financial accountants (J. L. Glick *Ann. NY Acad. Sci.* **265**, 178–192; 1976). Data audits are common for corporate biotechnology laboratories, but not for academic ones.

I estimated that the costs of funding questionable research practices (such as data misrepresentation and fabrication of results) could be reduced by US\$5–10 for each dollar spent on data audits (*Account. Res.* **2**, 153–168; 1992). Besides being a cost-effective way of monitoring the integrity of research organizations, data auditing helps to reveal genuine errors.

J. Leslie Glick Tampa, Florida, USA. *jlglick@ix.netcom.com*

FORUM Molecular biology

RNA discrimination

In the cell, genomic DNA is transcribed into various types of RNA. But not all RNAs are translated into proteins. Does this give protein-coding RNAs greater credibility in terms of function? Views differ.

THE TOPIC IN BRIEF

- Among RNA populations, messenger RNA arguably holds special status.
- Its mere transcription from DNA is considered sufficient evidence for its function as a protein-coding sequence.
- Whether tens of thousands of

non-protein-coding RNAs (ncRNAs) are equally important is debatable.

- One argument is that unless a function is discovered for a ncRNA, transcription per se is not enough to suggest that it has a function.
- The alternative viewpoint is that if ncRNAs are transcribed, it must be for a reason.

Quantity or quality?

MONIKA S. KOWALCZYK & DOUGLAS R. HIGGS

In many species, ncRNAs are abundant and bewilderingly complex. What we wonder is whether they all carry genetic information (as do all mRNAs), or whether some of them are the by-products of abnormal or inconsequential transcription.

Many short ncRNAs, which are often derived from long ncRNAs (lncRNAs), regulate gene expression (Table 1). Moreover, full-length lncRNAs may themselves have biological roles. Take, for example, the RNA product of the *XIST* gene, an lncRNA that effects inactivation of the X chromosome. Many ncRNAs are transcribed from intergenic regions around genes and their regulatory elements (for example, enhancers and promoters). Some overlap with protein-coding genes in both sense — the direction of transcription — and antisense orientations. Others lie in intergenic regions far from protein-coding genes.

In contrast to genes, however, genomic sequences encoding ncRNAs have often been poorly conserved during evolution. Also, very few natural mutations in ncRNAs have been shown to be the main cause of genetic diseases in humans, and few functionally important mutations in ncRNA-encoding genes have been identified in animal models¹. This suggests that, in contrast to many protein-coding genes, individual ncRNAs have a relatively minor role in biological processes.

If some ncRNAs are non-functional, why are they transcribed? Here, it may be that the level of expression is the significant factor. The

ever-improving technologies for sequencing the transcriptome² (the cell's complement of total RNA) can now detect RNAs present at an average of less than one copy per cell. At what stage does such information pass from revealing additional genetic complexity to simply detecting the inevitable by-products of transcription from accessible, activated chromatin (DNA-protein complexes)?

Indeed, the efficiency with which all stages of transcription and RNA processing are performed is intimately related to the physical and chemical state of the associated chromatin. The cell has evolved complex systems to suppress promiscuous transcription from chromatin both within genes and from intergenic regions. Furthermore, specific pathways degrade aberrant or irrelevant transcripts³. When these constraints are removed experimentally (and presumably when naturally modified *in vivo*), some irrelevant RNA transcripts are likely to be produced from many promoter- and enhancer-like elements that are accessible in chromatin.

The genome contains many more enhancers than it does protein-coding genes, and these determine when and where genes are expressed. Enhancers are widely dispersed throughout the genome. Those that occur between gene sequences produce a variety of RNAs, including lncRNAs, but very few of these transcripts have been shown unequivocally to have a function (Table 1). Enhancers located within genes also produce long transcripts known as multi-exonic enhancer RNAs (meRNAs; Table 1), which resemble mRNA. Nonetheless, these

transcripts have a very low coding capacity⁴, and their role — if any — is unknown.

Although many RNAs emanating from enhancers, promoters and other genomic elements that regulate gene expression may represent inconsequential transcription, transcription per se may be required for establishing or maintaining the activity of these elements and for 'templating' the associated chromatin. If so, information carried in the sequences of these ncRNAs would be largely irrelevant.

The huge number and complexity of RNAs being documented is certainly of great interest, and it would be surprising if evolution had not selected a proportion of these for their biological function⁵. However, the onus is on scientists to unequivocally demonstrate the biological roles of these molecules, rather than presuming that they are all functionally relevant components of the transcriptome. For starters, ncRNAs should be accurately classified by fully defining their associated transcriptional units and their patterns of expression during development and cell differentiation, as recently set out⁶. This, in turn, should direct the challenging experiments required to determine how various ncRNAs act, individually or in groups, to exert their proposed biological effects.

Monika S. Kowalczyk and Douglas R. Higgs are at the MRC Molecular Haematology Unit, Weatherall Institute of Molecular Medicine, University of Oxford, Oxford OX3 9DS, UK. e-mail: doug.higgs@imm.ox.ac.uk

Patience is a virtue

THOMAS R. GINGERAS

An open mind is not an uncritical one, and the obligation to be critical as scientists should not necessarily condition us to look unfavourably on unexpected results. We should therefore not be too sceptical about ncRNAs just because we don't know their functions. But let me start with an outline of where the debate originates.

TABLE 1 | MAIN CLASSES AND FUNCTIONS OF MAMMALIAN NON-CODING RNAs

ncRNA*	No. of known transcripts†	Transcript lengths (nucleotides; nt)‡	Functions
Precursors to short RNAs			
miRNA	1,756	>1,000	Precursors to short (21–23 nt) regulatory RNAs
snoRNA	1,521	>100	Precursors to short (60–300 nt) RNAs that help to chemically modify other RNAs
snRNA	1,944	1,000	Precursors to short (150 nt) RNAs that assist in RNA splicing
piRNA	89	Unknown	Precursors to short (25–33 nt) RNAs that repress retrotransposition of repeat elements
tRNA	497	>100	Precursors to short (73–93 nt) transfer RNAs
Long ncRNAs			
Antisense ncRNA	5,446	100–>1,000	Mostly unknown, but some are involved in gene regulation through RNA interference
Enhancer ncRNA (eRNA)§	>2,000	>1,000	Unknown
Enhancer ncRNA (meRNA)	Not fully documented	As variable as the length of mRNAs	Unknown, but they resemble alternative gene transcripts
Intergenic ncRNA	6,742	10 ² –10 ⁵	Mostly unknown, but some are involved in gene regulation
Pseudogene ncRNA	680	10 ² –10 ⁴	Mostly unknown, but some are involved in regulation of miRNA
3' UTR ncRNA	12	>100	Unknown

*miRNA, microRNA; snoRNA, small nucleolar RNA; snRNA, small nuclear RNA; piRNA, piwi-interacting RNA; tRNA, transfer RNA; antisense ncRNA, transcripts mapping and overlapping coding and non-coding RNAs; enhancer ncRNA (eRNAs and meRNAs), transcripts that initiate within regions that regulate specific genes; intergenic ncRNA, transcripts that map to genome regions between annotated genes; pseudogene ncRNA, transcripts that come from processed or unprocessed pseudogenes; 3' UTR ncRNA, 3'-untranslated regions of ncRNA.

†From ref. 13.

‡Summarized from a range of lengths.

§From ref. 16. Transcript number listed comes from the analysis of one cell line (mouse neuronal cells) and is a significant underestimate.

||From ref. 4. Analysis was done in mouse erythroid cells.

Within the past decade, reports that the mouse and human genomes were pervasively transcribed (meaning “that the majority of its bases are associated with at least one primary transcript”⁷) into predominantly ncRNAs^{8,9} were surprising and resulted in two types of criticism. The first of these centred on whether the detected RNAs are artefacts of the technologies used to identify them. The second objection focused on the biological importance of such transcripts. Unfortunately, these criticisms were often intermingled¹⁰, requiring subsequent correction¹¹.

The overwhelming majority of novel ncRNAs have three properties that suggest they should be ignored. The transcripts seem to have greatly reduced protein-coding potential; their expression levels are markedly lower than those of mRNAs; and their expression is mostly cell-type specific. Moreover, genomic sequences encoding these transcripts map to regions that were previously thought to be either untranscribed (sequences in the opposite strands to genes, and sequences between genes) or uninformative (intron sequences within genes, which do not make it into mature mRNA).

The artefact objection has now been addressed by results from many labs showing a wealth of ncRNA expression using a wide range of technologies (tiling arrays, high-throughput RNA sequencing, full-length complementary DNA cloning, northern hybridization and RNase protection). Determining the biological importance of ncRNAs is more challenging and an area of active investigation. Nonetheless, as the efforts to catalogue and characterize such RNAs get

under way, the initial atmosphere of scepticism continues to hang over this subject.

Healthy scepticism is an essential element of the scientific process. But it seems curious that ncRNAs have been deemed less interesting than mRNAs simply as a result of the short time since their discovery and the poor understanding of their biological roles. It is perhaps worth recalling that it took almost eight years from the discovery of the first member of

“We should not be too sceptical about non-coding RNAs just because we don’t know their functions.”

microRNAs (lin-4 miRNA) to the elucidation of the function of the very large class of short ncRNAs to which it belongs¹². The functions attributed to miRNAs include such fundamental biological

processes as control of developmental timing (miR273), organ development (miR84), tissue growth (miR181) and tumorigenesis (miR17).

According to the careful annotation by the GENCODE group¹³, there are currently some 161,000 human transcripts, 85,323 (53%) of which are ncRNAs¹⁴. Although the biological function of most of these ncRNAs is unclear, roughly 2% are precursors to miRNAs. In addition, 10% are lncRNAs that map to intergenic and intronic regions, and many of these transcripts have been implicated in regulation — both locally and from a distance — of developmentally important genes⁶. Notably, another 16% of the annotated ncRNAs map to pseudogenes — genes that have lost their original functional abilities. And some of these have been shown to regulate gene

expression by acting as decoys for miRNAs¹⁵.

With the growing identification of functional classes of ncRNA and understanding of the various roles that many of these transcripts have, the original atmosphere of pessimism concerning their biological importance should gradually change to one of cautious interest. The scientific process is not free of bias. But openness to fresh possibilities has the potential to reveal many new ideas. ■ [SEE INSIGHT P.321](#)

Thomas R. Gingeras is at the Genome Research Center, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA.
e-mail: gingeras@cshl.edu

1. Mattick, J. S. *PLoS Genet.* **5**, e1000459 (2009).
2. Mercer, T. R. *et al. Nature Biotechnol.* **30**, 99–104 (2011).
3. Schmid, M. & Jensen, T. H. *Wiley Interdisc. Rev. RNA* **1**, 474–485 (2010).
4. Kowalczyk, M. S. *et al. Mol. Cell* <http://dx.doi.org/10.1016/j.molcel.2011.12.021> (2012).
5. van Bakel, H. & Hughes, T. R. *Briefings Funct. Genomics Proteomics* **8**, 424–436 (2009).
6. Ulitsky, I., Shkumatava, A., Jan, C. H., Sive, H. & Bartel, D. P. *Cell* **147**, 1537–1550 (2011).
7. The ENCODE Project Consortium *Nature* **447**, 799–816 (2007).
8. Okazaki, Y. *et al. Nature* **420**, 563–573 (2002).
9. Kapranov, P. *et al. Science* **296**, 916–919 (2002).
10. van Bakel, H., Nislow, C., Blencowe, B. J. & Hughes, T. R. *PLoS Biol.* **8**, e1000371 (2010).
11. Clark, M. B. *et al. PLoS Biol.* **9**, e1000625 (2011).
12. Lee, R. C., Feinbaum, R. L. & Ambros, V. *Cell* **75**, 843–854 (1993).
13. Harrow, J. *et al. Genome Biol.* **7**, Suppl. 1, S4 (2006).
14. www.genencodegenes.org/stats.html
15. Ebert, M. S. & Sharp, P. A. *Curr. Biol.* **20**, R858–R861 (2010).
16. Kim, T.-K. *et al. Nature* **465**, 182–187 (2010).

Ultrasensitive radiocarbon detection

Radiocarbon is rare, forming no more than one part per trillion of the total carbon content of the atmosphere. An optical method allows radiocarbon to be detected at roughly 25-fold lower levels than this, opening up fresh avenues of research.

RICHARD N. ZARE

Radiocarbon dating is an invaluable technique for determining the age of carbon-containing samples up to about 50,000 years old. Until now, the only method available for measuring levels of radiocarbon (carbon-14) in a sample has been high-energy accelerator mass spectrometry, but the apparatus involved is bulky, expensive and complex. Reporting in *Physical Review Letters*, Galli *et al.*¹ describe an optical technique for measuring radiocarbon concentration that might overcome these problems. Their approach promises to greatly extend the use of radiocarbon measurements for dating, and as a tracer technique for following the fate of organic compounds in the body. It forms part of a growing revolution that is replacing mass spectrometry with optical methods for isotope analysis^{2–4}.

On Earth, there are three naturally occurring isotopes of carbon. The most abundant of these (99%) is carbon-12, with most of the rest being carbon-13. But carbon-14 also occurs in trace amounts, forming only as much as 1 part per trillion (0.0000000001%) of the carbon in Earth's atmosphere. Unlike ¹²C and ¹³C, ¹⁴C is radioactive, decaying with a half-life of about 5,730 years — a fact that makes it potentially useful as a radiolabel for several applications.

Carbon-14 is mainly produced in Earth's atmosphere from the bombardment of nitrogen molecules by cosmic rays. Plants fix atmospheric carbon dioxide during photosynthesis, and so the level of ¹⁴C in plants (and in animals that eat plants) when they die approximately equals the level of the isotope in the atmosphere at that time. Because the amount of ¹⁴C in dead organisms subsequently decreases as a result of radioactive decay, the date of carbon fixation (or death) can be determined by measuring the amount of the isotope in the remains. This is the basis of radiocarbon dating, the technique that has been a workhorse for estimating the age of organic remains from archaeological sites. In practice, the ratio of the number of ¹⁴C atoms to the total number of other carbon atoms in a sample is measured.

Because of the paucity of ¹⁴C, radiocarbon dating presents a huge technical challenge.

Most isotope-ratio measurements are carried out using mass spectrometry, in which ions are weighed by measuring their trajectories in electric and/or magnetic fields in a vacuum. Standard mass spectrometers, however, do not have sufficient resolution to distinguish the small mass difference between ¹⁴C and ¹⁴N, the most common isotope of nitrogen. The abundant presence of ¹⁴N therefore tends to mask any signal from radiocarbon.

High-energy accelerator mass spectrometry can overcome this problem. For this technique, the sample must first be turned into solid carbon (graphite) using a series of chemical transformations. It is then bombarded with caesium ions to produce negatively charged carbon ions, which are accelerated by a positive voltage of millions of volts. The negative ions, by now travelling at a few per cent of the speed of light, are subsequently converted into positive ions by an electron stripper (which consists of a gas or a thin foil), before the ions' masses are determined. Because nitrogen atoms do not form stable negative ions, the resulting data are free from nitrogen interference. But although accelerator mass spectrometers are powerful tools, they are also costly. Establishing and maintaining such an instrument costs millions of dollars, and so they tend to be found only at national facilities.

A much simpler approach is to completely oxidize a sample so that every carbon atom is turned into carbon dioxide³. The various isotopic forms of carbon dioxide can then be distinguished from each other because each has a slightly different infrared spectrum (they absorb slightly different frequencies of infrared light). All that is required to determine the ratios of carbon isotopes in a sample of carbon dioxide is to precisely measure the intensities of the spectral lines that correspond to infrared absorption for each isotopic form of the gas. Other compounds, such as water vapour and nitrogen, do not interfere in the infrared spectrum, either because they have different infrared spectra or because they do not have 'allowed' infrared transitions — that is, quantum mechanics prevents the molecules from undergoing energy transitions that would be detected in the infrared.

This optical approach has already been used³ to determine the carbon-isotope ratio

of carbon dioxide containing ¹²C and ¹³C, but the low concentration of ¹⁴C has made its measurement in carbon dioxide extremely difficult. Using an ultrasensitive technique called saturated-absorption cavity ring-down spectroscopy⁵, Galli *et al.*¹ have now succeeded in measuring the ratio of ¹⁴C to total carbon at values well below radiocarbon's natural abundance in carbon dioxide.

In their technique, the authors placed a gas sample between two or more highly reflecting mirrors that form an optical cavity. Infrared light that is incident on the cavity continually circulates within it, so that it takes many round trips. This effectively increases the optical path length of the light, allowing infrared absorption by the gas to be detected with a sensitivity that vastly exceeds what can be achieved in traditional absorption experiments.

Another feature of the cavity is that, when the infrared light source is interrupted, the radiant energy stored in the cavity 'rings down' — it decreases over time. Using a powerful infrared laser to 'saturate' the vibrational-rotational transitions in carbon dioxide that correspond to infrared absorption, Galli *et al.* used the rate of ring down as an excellent absolute measure of the concentration of absorbers inside the cavity (an approach that has previously been reported for infrared spectroscopy⁵). The authors obtained a linear concentration response down to a detection limit of about 43 parts per quadrillion, which makes their technique quite well suited for radiodating carbonaceous samples. It may also have applications in positron emission tomography (an imaging technique used in medicine for body scans), which often requires⁶ monitoring of carbon dioxide labelled with carbon-11, an artificial radioactive isotope of carbon.

Galli and colleagues say that the size of their experimental set-up is roughly two square metres in area, about 100 times smaller than the footprint of typical accelerator mass spectrometers. Furthermore, the equipment costs only about US\$400,000 — many times less than an accelerator mass spectrometer. For widespread adoption of the infrared technology, however, it will be necessary to reduce the cost even further, say by a factor of five or ten. Even so, an infrared method for measuring isotope ratios represents a real breakthrough because of the many possible uses of the technique. And there are other advantages. For example, in mass spectrometry, an ion from a sample is counted only once because its measurement neutralizes it. But infrared-absorption measurements do not destroy the sample, allowing it to be repeatedly analysed.

With further improvements, the infrared technique may well become the method of choice for measuring the isotope ratios of many common elements. Moreover, if the anticipated cost reductions are realized, the measurement of isotope ratios might become

a widely used tool in determining the origins of materials used for a broad range of purposes, from environmental monitoring to medical research. ■

Richard N. Zare is in the Department of Chemistry, Stanford University, Stanford, California 94305-5080, USA.
e-mail: zare@stanford.edu

BIODIVERSITY

Species choked and blended

The appearance of new ecological niches propels the evolution of species, but the converse can also occur. A study shows that changing lake habitats have caused extinctions and reduced the genetic differences between species. [SEE ARTICLE P.357](#)

JEFFREY S. MCKINNON & ERIC B. TAYLOR

Conventional wisdom long held that even if individuals of two different species could mate with each other, their offspring were doomed to early death or sterility. But a different view is taking hold: that it is often adaptations to different environments that cause species to separate, such that hybrid offspring fail because of their poor fit to resources, rather than through intrinsic shortcomings¹. As a consequence, changes to particular environmental conditions that previously kept species distinct could increase genetic mixing, and thereby reduce species number. On page 357 of this issue, Vonlanthen *et al.*² provide evidence that human alterations

to lake habitats have eroded barriers between species and contributed to extinctions.

The authors' study of 17 Swiss lakes shows that glacial melting in the past 12,000 years provided ecological opportunities, in the form of new environmental niches, that led to diversification of whitefish species, as has been reported for other freshwater fishes³. Whitefish species divergence is characterized by, for example, differences in body size and the number of 'gill rakers' — cartilaginous structures that protrude from fish gills and are involved in feeding (Fig. 1). Large-bodied whitefish, which have fewer gill rakers, typically feed from the bottom of lakes and spawn in shallow water in winter, whereas smaller species, which have more gill rakers, tend to

feed in open water and spawn much deeper.

However, increased human activity around the lakes dramatically altered the lakes' ecology during the twentieth century. Higher nutrient levels in the water caused eutrophication, in which algal populations increase, water quality is reduced and oxygen levels at the lake bottom decrease. Vonlanthen *et al.*¹ propose that these conditions compressed the depth range in which whitefish could spawn, bringing previously separated species together to breed, forming hybrids. Whitefish feeding patterns were probably also affected, through reductions in zooplankton diversity and possibly in the density of bottom-dwelling prey (Fig. 1), which would also have reduced opportunities for exploiting ecological variation.

Vonlanthen and colleagues' data show that the extent of species loss for each lake correlates with the severity of that lake's eutrophication. But did these extinctions result exclusively from demographic decline — the extinction process we usually think of, in which deaths outnumber births? Or was reverse speciation at play, in which characteristics that once defined distinct species are merged into a single hybrid species?

The authors report² several lines of evidence suggesting a role for reverse speciation in the lakes. First, the severity of eutrophication is the best predictor of genetic differentiation of modern whitefish — lakes that suffered the greatest eutrophication contain species that are the least genetically different from each other. Historical DNA samples also allowed Vonlanthen and colleagues to document a progressive reduction in whitefish genetic differentiation in one of the lakes (Lake Constance) between 1926 and 2004. Furthermore, they find strong genetic traces of the extinct whitefish species

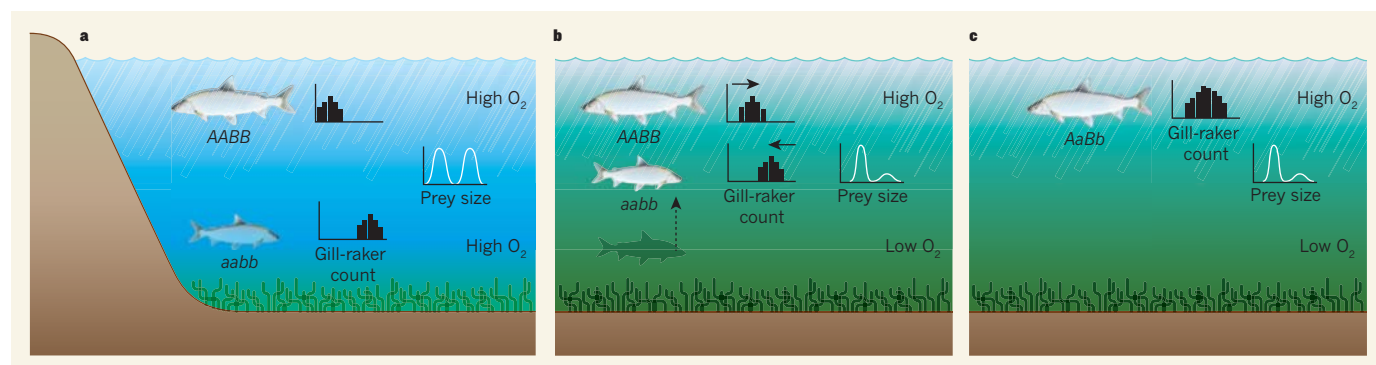


Figure 1 | Loss of fish biodiversity through eutrophication. **a**, Before human activity raised nutrient levels in lake waters, the Swiss lakes studied by Vonlanthen *et al.*² were well oxygenated at all depths, and there were diverse invertebrate prey communities in both the open water (suggested by other studies to be generally smaller prey, represented by the left side of the prey-size distribution) and at the bottom (generally larger prey, right side of distribution). These resources supported genetically distinct species of whitefish (represented as AABB and aabb) with different characteristics, including their body size and number of gill rakers — cartilaginous protrusions from the gills. Large-bodied whitefish with fewer gill rakers generally fed from the bottom and spawned in

shallow water, whereas small-bodied species with more gill rakers typically fed in open water and spawned much deeper. **b**, Lake eutrophication led to lower oxygen levels, especially at depth, driving deep-spawning species into shallower water, where they spawned with other species to form hybrids. Simultaneously, the fishes' prey became less diverse, thereby reducing divergent selection — the process by which different ecological niches provide a selective pressure for species to have distinct characteristics. **c**, Increased hybridization and reduced divergent selection, as well as demographic decline, resulted in extinction of the deeper-spawning species, with the remaining species being a genetic hybrid (AaBb) and possessing an intermediate number of gill rakers.

Coregonus gutturosus in extant sister species, implicating hybridization in that extinction. The authors also document lessened differences in the fishes' gill-raker numbers, a key characteristic, in the most polluted lakes. This finding is consistent with the hypothesis that eutrophication reduced ecological opportunity, which in turn weakened selection for differences in feeding traits.

Previous cases of reverse speciation in fishes^{4,5} and birds⁶ have shown that altered ecological conditions^{7,8} can erode fragile reproductive barriers and allow the formation of viable hybrids. However, the mechanisms of species collapse have often remained obscure. The current study is noteworthy because it establishes strong links among changed environmental conditions, reduced ecological opportunity and reverse speciation. The scale of the effect in whitefish, studied over decades and across 17 lakes, is also exceptional. The work highlights an under-appreciated aspect of biodiversity loss — 'cryptic extinction', whereby considerable morphological and genetic variability is maintained within hybrids, but previously species-specific combinations of these features are lost.

Cryptic extinction may have a particularly high impact on fish biodiversity because individual lakes often contain unique species, and fresh waters contain about 40% of all fish species⁹. But reverse speciation can also occur in terrestrial environments, particularly those similar to lakes, such as volcanic islands⁶.

The major limitation of Vonlanthen and colleagues' study is its correlational nature. Whitefish hybridization clearly increased in the Swiss lakes as pollution and disturbance increased, but factors in addition to those highlighted by the authors may have contributed to the loss of diversity. One of these is a by-product of demographic decline — as one species becomes rare, finding mates becomes more difficult, and so more frequent hybridization would be expected. Other potential confounding processes include the introduction of whitefish from hatcheries, overfishing and the impact of invasive species. However, despite these other influences, a convincing effect of eutrophication levels on biodiversity emerges from the study².

The work raises a number of additional important questions. How much, and which parts, of the genomes of extant whitefish species are 'original' compared with hybrid in origin? Which genes are responsible for the critical differences between whitefish species, and how has the prevalence of variants of these genes altered in response to ecological changes? In addition, what are the relative roles of the two processes of increased hybridization and reduced divergent selection (in which the existence of multiple ecological niches promotes the divergence of distinct species) in driving reverse speciation? Genome-wide analyses of both historical and modern whitefish samples

will help to address these questions.

A more practical concern is what happens next. Eutrophication has now been eliminated or greatly reduced in most of the lakes studied, so they more closely resemble their previous state. Can we expect 're-speciation', in which fishes with characteristics of extinct species reappear¹⁰? Current theory does not provide a clear answer, but suggests that distinct species can re-emerge after a brief collapse¹¹. If Vonlanthen and colleagues are correct and speciation reversal is an under-appreciated threat to biodiversity, we need to understand how to prevent and correct the ecological changes responsible — and perhaps learn how to recognize when it truly is too late. ■

Jeffrey S. McKinnon is in the Department of Biology and North Carolina Center for Biodiversity, East Carolina University, Greenville, North Carolina 27858, USA.

Eric B. Taylor is in the Department of Zoology, Beaty Biodiversity Research Centre and Museum, and the Native Fishes Research Group, University of British Columbia, British Columbia V6T 1Z4, Canada.
e-mails: mckinnonj@ecu.edu;
etaylor@zoology.ubc.ca

1. Schluter, D. *Science* **323**, 737–741 (2009).
2. Vonlanthen, P. et al. *Nature* **482**, 357–362 (2012).
3. Schluter, D. *The Ecology of Adaptive Radiation* (Oxford Univ. Press, 2000).
4. Seehausen, O., van Alphen, J. J. M. & Witte, F. *Science* **277**, 1808–1811 (1997).
5. Taylor, E. B. et al. *Mol. Ecol.* **15**, 343–355 (2006).
6. De León, L. F. et al. *Evolution* **65**, 2258–2272 (2011).
7. Behm, J. E., Ives, A. R. & Boughman, J. W. *Am. Nat.* **175**, 11–26 (2010).
8. Rhymer, J. M. & Simberloff, D. *Annu. Rev. Ecol. Syst.* **27**, 83–109 (1996).
9. Dudgeon, D. et al. *Biol. Rev.* **81**, 163–182 (2006).
10. Turner, G. F. *Fish Fisheries* **3**, 225–229 (2002).
11. Gilman, R. T. & Behm, J. E. *Evolution* **65**, 2592–2605 (2011).

EARTH SCIENCE

Intraplate volcanism

The origin of volcanic activity occurring far from tectonic-plate boundaries has been a subject of contention. The latest geodynamic model offers a fresh take on the matter. SEE LETTER P.386

CIN-TY A. LEE & STEPHEN P. GRAND

In this issue, Liu and Stegman¹ present a hypothesis for the generation of volcanic centres that might change our view of how plate tectonics influences the distribution of volcanic activity on Earth.

The theory of plate tectonics describes the uppermost of Earth's layers as made up of rigid plates, the relative motions of which are confined to narrow plate boundaries. The boundaries come in three types: divergent, where plates move away from one another and create systems such as mid-ocean ridges; convergent, where one plate slides beneath another, forming subduction zones; and transform margins, where plates slide past one another, as in the San Andreas Fault system.

Plate tectonics successfully explains most of Earth's geological features. For example, volcanism at mid-ocean ridges can be explained by decompression melting associated with passive upwelling of hot (asthenospheric) mantle in response to plate divergence. Volcanism at subduction zones can be described by a combination of two effects: partial melting of the mantle, driven by return flow in the mantle wedge overlying the subduction zone, and melting-point depression, caused by the influx of water released from the descending plate of the subduction zone.

Volcanoes that occur far from plate

boundaries — for example, intraplate magmatism — are more difficult to explain with plate tectonics. Some intraplate volcanic systems, such as the Hawaiian volcanic chain in the Pacific plate and the Yellowstone volcanic field in North America, migrate along tracks that seem independent of plate-boundary processes. The effusive but short-lived outpourings of basalts, known as flood basalts, some of which are so large that they cover substantial areas of continents or even entire plates, are also not easily described by the interaction of slowly moving plates.

One popular view is that intraplate magmatism is driven by narrow mantle upwellings (plumes) originating from a hot thermal layer at the core-mantle boundary². Therefore, the expression of plumes at Earth's surface should be independent of plate motions². Flood basalts are thought to record the initial impingement of the anomalously hot plume head, whereas the volcanic track, known as the hot-spot track, records the passage of the plate over the plume's tail³. For example, the eruption of the Steens-Columbia River flood basalt about 17 million years ago is thought to represent the initiation of the currently active Yellowstone hot-spot track, and so is conjectured to fit into the plume theory^{4,5}.

However, the eruption area of the Steens-Columbia River flood basalt is oriented north-south, perpendicular to the

Yellowstone track. In addition, the geochemistry of the flood basalt differs from that of the Yellowstone volcanics^{6,7}, complicating the plume hypothesis. Alternatively, the Steens–Columbia River flood basalt could be associated with extension of the upper plate behind the Cascades volcanic arc⁸ (back-arc spreading). But this phenomenon does not seem to explain the sudden appearance of the Steens–Columbia River flood basalt.

In their study, Liu and Stegman¹ (page 386) propose that the Steens–Columbia River flood basalt is a natural consequence of slowing convergence between the North American plate and the ancient Farallon plate. This slow-down was presumably associated with the approach of a mid-ocean ridge between the Farallon and Pacific plates 20 million years ago, now manifested as the active Juan de Fuca ridge. The authors performed geodynamic calculations with initial and boundary conditions constrained by observed relative plate motions and plate age. They show that stretching and eventual tearing of the Farallon plate accompanied the slow-down of convergence, resulting in detachment of the Farallon plate.

Liu and Stegman find that the model that best reproduces the presumed current location of the Farallon plate, as constrained from seismic tomography, predicts tearing to have begun about 16 million to 17 million years ago, when the Steens–Columbia River flood basalt initiated. Dynamic pressures generated from this tear resulted in rapid mantle upwelling through this gap in the slab, driving a magmatic flare-up that mimics the structural trend of the Steens–Columbia River flood basalt (Fig. 1).

If Liu and Stegman's model is correct, the implication is that some intraplate magmatism can be explained by the development of gravitational instabilities within subducting slabs. In their model, thermal upwelling is still responsible for flood basalts, but unlike traditional plumes, which derive from the lowermost mantle, an upper-mantle origin is implied. There are, however, some features that remain unresolved. For example, the model does not provide a good explanation for the eastward migration of the Yellowstone hot-spot track, the high ratio of helium-3 to helium-4 in Yellowstone volcanics⁷ or the presence of a seismic-velocity anomaly extending into the lower mantle beneath Yellowstone⁹. And it may not explain the isotopic signatures seen in the Steens–Columbia River flood basalt.

If accurate, Liu and Stegman's model should apply to other locations where slab tears have occurred. Such a tear clearly happened in central California about 20 million years ago, because the last remnants of the Farallon plate were captured on the coast of California, but the rest of the Farallon is no longer present beneath the state¹⁰. There is evidence of a flare-up in basaltic magmatism east of central California, for example on the border of California



Figure 1 | Columbia River Gorge, Oregon. Large, effusive outpourings of basalts, such as the Steens–Columbia River flood basalt in Oregon and Washington exposed here on the margins of the river, are usually attributed to the impingement of thermal plumes arising from the core–mantle boundary. Liu and Stegman show¹ that the timing and distribution of eruption may instead be related to tears developed within subducting slabs.

and Nevada, during this time. But the magnitude does not seem comparable to that of the Steens–Columbia River flood basalt, suggesting that different boundary conditions might need to be considered in the authors' model.

In any case, Liu and Stegman's study is pertinent because it draws more attention to subducting slabs in generating intraplate magmas. The following examples might be considered. Where a young oceanic plate is subducting, a slab tear, accompanied by large-volume magmatic flare-ups, should develop because young plates are difficult to subduct. This hypothesis may apply to the eastern Pacific. By contrast, when an old oceanic plate is subducting, a long segment of the slab might be expected to stagnate temporarily in the transition zone between the upper and lower mantle¹¹. The juxtaposition of cold slab material against hot mantle at depth would generate small-scale thermal upwellings along the edges of the slab¹². These upwellings could generate widespread basaltic magmatism far from the subduction trench, as seen in northeastern China. If the edges of the slab are migrating relative to the upper plate, hot-spot tracks could be generated¹³.

We note that all of these upwellings are sourced in the upper mantle and likely to be superimposed on magmatism associated with back-arc spreading; thus a complicated pattern of magmatism is expected. Should a subducting slab penetrate deep into the lower mantle, upwellings might be expected even further from plate boundaries.

In conclusion, there is reason to speculate that intraplate magmas might be intimately linked to subducting slabs^{12,14}. In other words, it is conceivable that plate tectonics generates many intraplate magmas. Differences in the magnitude and locations of intraplate

magmas may simply be controlled by the scale of subducting slabs. The debate over whether deep-seated thermal plumes exist¹⁵ remains unresolved because these narrow upwellings are difficult to detect. An alternative approach is to map out the geometry and length scale of subducting slabs, which may be easier to detect by various geophysical methods. Liu and Stegman's model shows how downwellings, such as subduction, must be considered when understanding the origin of upwellings and their associated magmatic activities. ■

Cin-Ty A. Lee is in the Department of Earth Science, Rice University, Houston, Texas 77005, USA. **Stephen P. Grand** is in the Department of Geological Sciences, University of Texas at Austin, Texas 78749, USA. e-mail: ctlee@rice.edu

1. Liu, L. & Stegman, D. R. *Nature* **482**, 386–389 (2012).
2. Morgan, W. J. *Nature* **230**, 42–43 (1971).
3. Richards, M. A., Jones, D. L., Duncan, R. A. & DePaolo, D. J. *Science* **254**, 263–267 (1991).
4. Smith, R. B. et al. *J. Volcanol. Geotherm. Res.* **188**, 26–56 (2009).
5. Hooper, P. R., Camp, V. E., Reidel, S. P. & Ross, M. E. *Geol. Soc. Am. Spec. Pap.* **430**, 635–668 (2007).
6. Leeman, W. P., Schutt, D. L. & Hughes, S. S. *J. Volcanol. Geotherm. Res.* **188**, 57–67 (2009).
7. Graham, D. W. et al. *J. Volcanol. Geotherm. Res.* **188**, 128–140 (2009).
8. Carlson, R. W. & Hart, W. K. J. *Geophys. Res.* **92**, 6191–6206 (1987).
9. Schmandt, B. & Humphreys, E. *Earth Planet. Sci. Lett.* **297**, 435–445 (2010).
10. Nicholson, C., Sorlien, C. C., Atwater, T. A., Crowell, J. C. & Luyendyk, B. P. *Geology* **22**, 491–495 (1994).
11. Zhao, D. *Phys. Earth Planet. Inter.* **146**, 3–34 (2004).
12. Faccenna, C. et al. *Earth Planet. Sci. Lett.* **299**, 54–68 (2010).
13. James, D. E., Fouch, M. J., Carlson, R. W. & Roth, J. B. *Earth Planet. Sci. Lett.* **311**, 124–135 (2011).
14. James, D. E., Fouch, M. J., VanDecar, J. C. & van der Lee, S. *Geophys. Res. Lett.* **28**, 2485–2488 (2001).
15. Anderson, D. L. *Geol. Soc. Am. Spec. Pap.* **388**, 31–54 (2005).

A sweet way of sensing danger

Cells can destroy invading bacteria through a digestive process called autophagy. A study finds that sugar molecules, exposed by bacterial damage to the cell's membrane, can trigger this process. [SEE LETTER P.414](#)

JU HUANG & JOHN H. BRUMELL

The bacterium *Salmonella enterica* serovar Typhimurium is a leading cause of food poisoning and a well-studied model pathogen. On ingestion by its host, the microbe can penetrate and grow in the gut's epithelial cells, and damage cell membranes. These events can lead to inflammation and to the pathogen's spread to other tissues. To combat the infection, one defence mechanism available to cells is autophagy — a process by which the intracellular bacteria are digested in cytoplasmic vesicles known as lysosomes. On page 414 of this issue, Thurston *et al.*¹ report that sugar molecules normally present on the cell's surface can act as a 'danger' signal in the cytoplasm to target *S. Typhimurium* for autophagy.

Autophagy is a highly regulated process, during which cytoplasmic cargoes are captured in a double-membrane vesicle, the autophagosome. This vesicle then fuses with lysosomes, which are loaded with digestive enzymes that eventually degrade the vesicle's contents. Cells use autophagy to maintain a balance between the synthesis and degradation of their own components, and malfunction of the process has been linked to human diseases such as cancer, neurodegenerative disorders, diabetes and inflammatory bowel disease. The autophagy system can also target invading pathogens such as *S. Typhimurium* for degradation, although the precise mechanisms underlying this process are not well understood.

Early after cell invasion, *S. Typhimurium* resides in a vesicle known as the *Salmonella*-containing vacuole (SCV). Some of the bacteria then use specialized protein machinery (called a type III secretion system)² to generate pores in the SCV membrane, through which they deliver protein effectors into the cell's cytoplasm. The effectors modulate the activity of the host's cellular machinery to promote intracellular growth of the pathogen. Moreover, damaged SCVs allow some of the bacteria to escape and replicate in the cytosol. However, the microbes in damaged vesicles can also 'attract' components of the autophagy machinery, such as the protein LC3, as well as autophagosomes that engulf the damaged SCVs and restrict bacterial growth^{2,3} (Fig. 1).

So how does the autophagy system recognize

the bacteria in damaged SCVs? Host factors such as reactive oxygen species⁴ and the lipid diacylglycerol⁵ play a part in the recruitment of LC3 to SCVs. In addition, the microbes attract the protein ubiquitin, which seems to recruit autophagy 'adaptor' proteins that, in turn, bind to LC3. In particular, the adaptor proteins p62, NDP52 and optineurin contribute to LC3 recruitment to the bacteria^{6–8}. However, these adaptors are not functionally redundant — depletion of any of them impairs antibacterial autophagy — and they bind to different regions, or microdomains, around an individual bacterium⁹. These observations suggest that there may be different signature molecules on the microbial surface and/or on the damaged SCV that attract different adaptor proteins, and that SCV damage is essential for exposing these signature molecules.

To explore potential mechanisms of adaptor recruitment, Thurston *et al.*¹ focused on galectins, a family of proteins that bind complex sugar molecules. The authors found that, 1 hour after invasion by *S. Typhimurium*, galectins were associated with a small population of the bacteria in the host cells. When they abolished the expression of a specific galectin, galectin 8, intracellular bacteria grew at an increased rate. Furthermore, galectin 8 co-localized with NDP52 on the intracellular microbes, although at microdomains distinct from those to which p62 or ubiquitin bind.

Thurston and colleagues carried out experiments with live cells and purified proteins that confirmed a direct interaction between galectin 8 and NDP52. Moreover, the authors observed that cells lacking galectin 8 failed to recruit NDP52 to the bacteria at 1 hour after invasion, suggesting that galectin 8 acts as an early signal to attract autophagy adaptors. The recruitment of LC3 to the intracellular microbes was also significantly reduced. When the authors artificially targeted NDP52 to SCVs in the galectin-8-depleted cells, bacterial growth was again restricted. This indicates that galectin-8 recruitment of NDP52 is essential for the induction of antibacterial autophagy.

But how do the intracellular bacteria attract the sugar-binding galectin? SCV damage exposes the cytoplasm to different sugar molecules. Although some of these derive from the bacteria, there are also host molecules that

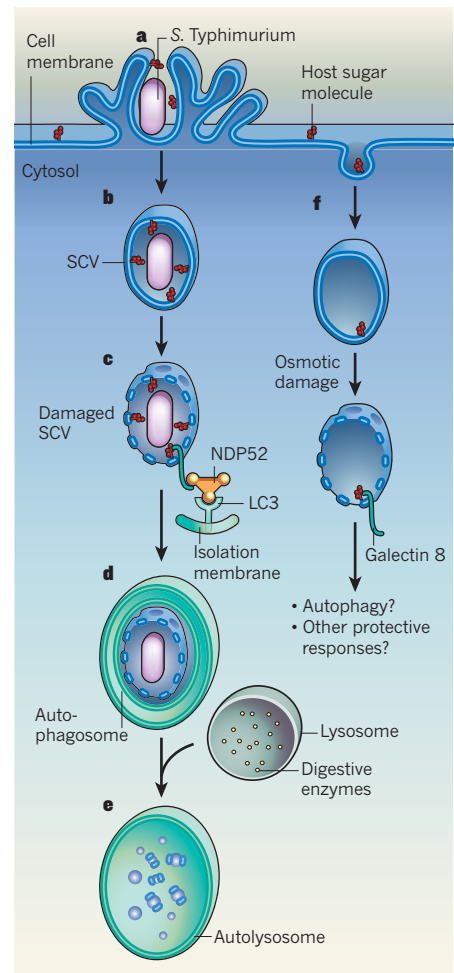


Figure 1 | Broken vesicles as danger signals.

a, b, Early after the pathogenic bacterium *S. Typhimurium* invades host cells (**a**), it resides in a cytoplasmic vesicle known as a *Salmonella*-containing vacuole (SCV; **b**). **c,** The bacterium can damage the vacuole membrane, exposing host sugar molecules to the cytoplasm. Thurston *et al.*¹ report that the host protein galectin 8 binds to these sugars and triggers a process called autophagy, by which the invading bacteria are destroyed. Specifically, the authors found that galectin 8 recruits another protein, NDP52, by direct interaction. NDP52, in turn, binds to the protein LC3 and recruits other components of the autophagy machinery to the damaged SCV. **d,** Eventually, the bacterium is enclosed in a specialized vesicle known as an autophagosome, which forms from the isolation membrane. **e,** Other vesicles (lysosomes) containing digestive enzymes can then fuse with the autophagosome, forming the autolysosome within which the pathogen is destroyed. **f,** The authors found that osmotic damage to cytoplasmic vesicles, in the absence of bacteria, also exposes host sugars that recruit galectin 8.

were initially present in the vesicle. Thurston *et al.*¹ report that purified galectin 8 does not bind to *S. Typhimurium in vitro*, which suggests that the molecule targeted by galectin 8 is not derived from the microbe. Furthermore, the targeting of galectin 8 to intracellular *S. Typhimurium* was severely impaired in host cells lacking complex sugar molecules, which

are normally found on the cell's surface and lining the inside of some vesicles. The authors found that other vesicle-damaging pathogenic bacteria such as *Listeria monocytogenes* and *Shigella flexneri* are also decorated by galectins soon after infection. Moreover, osmotic damage of cytoplasmic vesicles in the absence of bacteria also resulted in the recruitment of galectins to the damaged vesicles. On the basis of their observations, the researchers conclude that galectins can act as sensors of non-specific danger by detecting host sugar molecules that are exposed on damaged vesicle membranes.

Although it has been speculated² that damaged SCVs serve as a signal to target bacteria for autophagy, Thurston and colleagues' work provides much-needed insight into the mechanistic details. Their results show that, when the microbes try to escape into the cytoplasm by disrupting the vesicles, host sugar molecules are exposed. Cytoplasmic galectin 8 then functions as a danger receptor: it binds to the exposed carbohydrates and recruits NDP52, which further attracts LC3 and autophagy machinery to the damaged compartment, thus triggering antibacterial autophagy soon after infection.

But the authors also show that recruitment of galectin 8 to damaged vesicles is a general danger response. Whether autophagy is also activated by galectin 8 in any other situations in which a cellular organelle is disrupted needs to be further investigated. Other galectins are recruited to damaged cytoplasmic vesicles such as SCVs, but at present their role in cellular defences to infection is unclear.

Is sugar exposure the only signal required to detect damaged SCVs? Most likely not, because NDP52 is only one of the three adaptors required to target S. Typhimurium to autophagy. So, the mechanisms that regulate recruitment of p62 and optineurin to damaged SCVs, and the ways by which the three adaptors cooperatively regulate antimicrobial autophagy, are exciting questions for future studies. ■

Ju Huang and John H. Brumell are in the Cell Biology Program, Hospital for Sick Children, Toronto, Ontario M5G1X8, Canada. **J.H.B.** is also in the Department of Molecular Genetics and Institute of Medical Science, University of Toronto. e-mails: juh@sickkids.ca; john.brumell@sickkids.ca

1. Thurston, T. L. M., Wandel, M. P., von Muhlinen, N., Foeglein, A. & Randow, F. *Nature* **482**, 414–418 (2012).
2. Birmingham, C. L. *et al.* *J. Biol. Chem.* **281**, 11374–11383 (2006).
3. Kageyama, S. *et al.* *Mol. Biol. Cell.* **22**, 2290–2300 (2011).
4. Huang, J. *et al.* *Proc. Natl Acad. Sci. USA* **106**, 6226–6231 (2009).
5. Shahnazari, S. *et al.* *Cell Host Microbe* **8**, 137–146 (2010).
6. Zheng, Y. T. *et al.* *J. Immunol.* **183**, 5909–5916 (2009).
7. Thurston, T. L. M. *et al.* *Nature Immunol.* **10**, 1215–1221 (2009).
8. Wild, P. *et al.* *Science* **333**, 228–233 (2011).
9. Cemama, M., Kim, P. K. & Brumell, J. H. *Autophagy* **7**, 341–345 (2011).

ASTROPHYSICS

Echoes from an old outburst

Almost two centuries after the eruption of one of the most massive binary systems in our Galaxy, light reflected from its surroundings has been detected. The observations challenge traditional models for the eruption. SEE LETTER P.375

NOAM SOKER & AMIT KASHI

Extremely massive stars — at least a hundred times more massive than the Sun — are rare astrophysical objects. Historically, they were thought to influence their environment through their luminosity (equivalent to that of several million Suns) and through the explosion, or supernova, that marks the end of their life. On page 375 of this issue, Rest *et al.*¹ describe an analysis of light echoes from the nineteenth-century 'Great Eruption' of η Carinae, one of the most massive two-star systems in the Milky Way. The work strengthens previous claims that such stars have a third mechanism by which to release mass and energy into their environment.

From 1838 to 1858, η Carinae was in a continuous state of energetic outburst². At the time, however, there was no equipment available with which to record its spectrum. The spectrum of an astrophysical source, such as a star or galaxy, can be thought of as its fingerprints: it provides information about the source's temperature, density, velocity and chemical composition.

In an impressive piece of observational work, Rest and colleagues¹ return, almost two centuries later, to the 'crime scene' of η Carinae's eruption and take its fingerprints. The prime suspect in the case is the least massive star of the binary system, the companion, which stole mass from the more massive and more evolved 'primary' star. The primary itself was in an unstable phase during the 20-year period of the eruption. The crime scene is the ambient dust that reflects light from the eruption. This reflected light has taken longer to reach Earth than light following a straight path and not crossing the dust, and so has arrived there only recently — this is why the phenomenon is termed a light echo.

From spectral analyses of light echoes from η Carinae's eruption, Rest *et al.* deduce that the temperature of the gas ejected during the outburst was about 5,000 kelvin. This temperature, the authors say, is lower than that expected from conventional eruption models — in which the outflow of the eruption is in the form of a strong stellar wind. They further argue that such a temperature fits best with a hydrodynamic eruption mechanism. One possible model³ for a hydrodynamic eruption

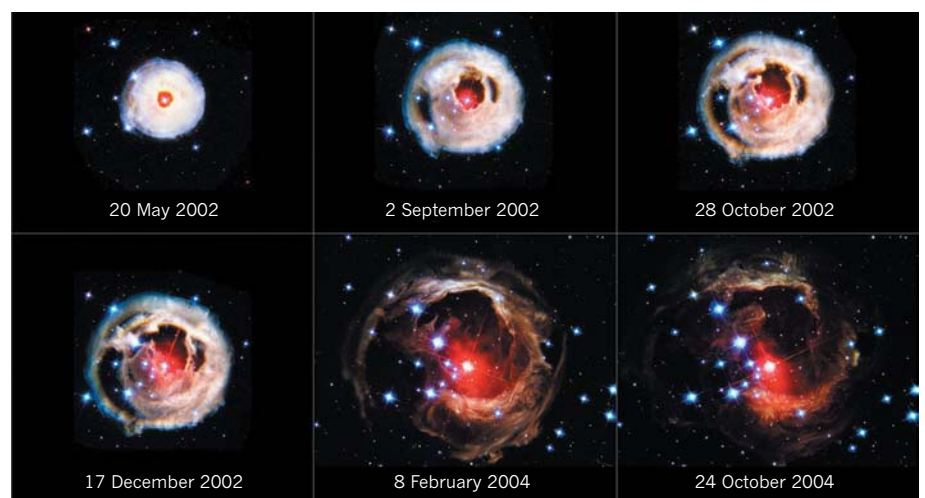


Figure 1 | V838 Monocerotis' light echoes. The six panels show a time sequence of light echoes from the 2002 eruption of the star V838 Monocerotis, as light from the eruption reached and illuminated the star's dusty surroundings, at increasing distances from the star, and travelled to Earth. The star itself is not resolved, but it is located at the centre of the structures in these images. The eruption of V838 Monocerotis and that of the nineteenth-century eruption of the two-star system η Carinae studied by Rest and colleagues¹ have some common properties, despite the different nature of the erupting stars.

NASA, ESA & Z. LEVAY (STSC)

involves mass transfer from the primary to the companion.

According to this binary model, during the 20-year Great Eruption, the companion would have accreted matter in the form of gas in an amount equivalent to several times the mass of the Sun. A huge amount of gravitational energy would have been released during this accretion process, which would have been the main energy source of the Great Eruption. Furthermore, some of the mass accreted by the companion would have been blown by the companion itself in two opposing directions, leading to the shaping of the Homunculus bipolar nebula, which is now observed to surround the binary system. Most of the mass in the nebula was blown directly by the primary star. The present masses of the primary and companion may be up to 170 and 80 times that of the Sun, respectively³.

During the Great Eruption, η Carinae experienced two bright peaks in luminosity, in 1838 and in 1843 (refs 4, 5). Rest *et al.*¹ find that the echoes' light curves — graphs of their intensity as a function of time — are consistent with these peaks. The time difference between the two peaks corresponds to the orbital period of the binary system around 1840; at present, the orbital period is five and a half years^{3,4,6}. The peaks themselves occurred when the two stars were closest together in their elliptical orbit around each other. Rest and colleagues' analysis of the echoes' spectra and light curves lends some support to an eruption model in which energy comes from mass transfer that is triggered at the stars' closest approach.

The temperature of about 5,000 K and the occurrence of two strong peaks (two weaker peaks are recorded historically at around 1849 and 1854) are reminiscent of the eruptive event⁷ that the star V838 Monocerotis experienced in 2002 (Fig. 1). One popular model for this eruption posits⁸ that a low-mass star of about half the mass of the Sun was destroyed in a merger with a star about six times more massive than the Sun. The accretion of gas from the low-mass star onto the surface of the more massive star would have been the energy source of the eruption. As in the case of η Carinae, the star that accreted mass is a non-evolved star such as the Sun: it is at an evolutionary stage during which nuclear-fusion reactions of hydrogen still occur in its centre.

The progenitor of η Carinae's eruption seems to fall into a varied group of systems that undergo eruptions powered by impulsive mass accretion onto non-evolved stars. The accreting stars can be very massive, as for η Carinae; five to eight times as massive as the Sun, as for V838 Monocerotis; or Sun-like stars. This heterogeneous group of progenitors might also include dying red-giant stars. Accretion of mass from a dying red-giant star onto a Sun-like star over a time span of 5–50 years could

lead to eruptions and shape some bipolar planetary nebulae. Red-giant stars are Sun-like stars in a late phase of evolution, during which they become very bright and large. Planetary nebulae are the last moment of a Sun-like star's glory: they are beautiful shining clouds of gas and dust that last for 100,000 years. The nebulae are formed from gas that was once part of the outer shells of the red-giant star. Some of these planetary nebulae are known to have been formed over a short period of time, and have a structure that is not unlike that of η Carinae. One example of such nebulae is the bipolar planetary nebula NGC 6302 (ref. 9).

As Rest and colleagues¹ mention, a few more years of data are required to improve the echoes' light curves and to test their consistency with the historical observations. This will definitely help to nail down the origin of the eruption event, to find out whether it was triggered by mass transfer to the companion or by some as-yet-undetermined eruptive event in the primary itself, as proposed by some traditional models. Although it has been studied for more than a century, η Carinae still holds

several secrets. In the coming years, it is hoped that observations with modern telescopes will shed more light on this intriguing binary system. ■

Noam Soker is in the Department of Physics, Technion, Haifa 32000, Israel. **Amit Kashi** is in the Department of Physics and Astronomy, University of Nevada, Las Vegas, Las Vegas, Nevada 89154–4002, USA. e-mails: soker@physics.technion.ac.il; kashia@physics.unlv.edu

1. Rest, A. *et al.* *Nature* **482**, 375–378 (2012).
2. Davidson, K. & Humphreys, R. M. *Annu. Rev. Astron. Astrophys.* **35**, 1–32 (1997).
3. Kashi, A. & Soker, N. *Astrophys. J.* **723**, 602–611 (2010).
4. Smith, N. & Frew, D. J. *Mon. Not. R. Astron. Soc.* **415**, 2009–2019 (2011).
5. Humphreys, R. M., Davidson, K. & Smith, N. *Publ. Astron. Soc. Pacif.* **111**, 1124–1131 (1999).
6. Damineli, A. *Astrophys. J.* **460**, L49–L52 (1996).
7. Bond, H. E. *et al.* *Nature* **422**, 405–408 (2003).
8. Soker, N. & Tylenda, R. *Mon. Not. R. Astron. Soc.* **373**, 733–738 (2006).
9. Szyszka, C., Zijlstra, A. A. & Walsh, J. R. *Mon. Not. R. Astron. Soc.* **416**, 715–726 (2011).

STRUCTURAL BIOLOGY

Ion channel in the spotlight

When expressed in neurons, channelrhodopsin proteins allow the cells' electrical activity to be controlled by light. The structure of one such protein will guide efforts to make better tools for controlling neurons. SEE ARTICLE P.369

OLIVER P. ERNST & THOMAS P. SAKMAR

Imagine taking a pigment from the eyespot (the light-receptive organelle) of a motile, photosynthetic alga and putting it into the neuron of a living mouse. Now imagine exciting the pigment using laser light and seeing a reproducible effect of this stimulus on the behaviour of the mouse. It sounds unbelievable, but this is the basis of optogenetics — the combination of optical techniques and genetic engineering that allows light to control an organism's physiology and behaviour¹.

The algal eyespot pigments that facilitate optogenetics are proteins called channelrhodopsins (ChRs), and they can be thought of as light-activated, nanometre-scale electrodes. When expressed in cells *in vitro* or *in vivo*, ChRs target the cell membrane and are bound to a chromophore — a kind of molecular antenna that absorbs light. Illumination of the ChR rapidly causes a flow of cations across the membrane. The resulting electrical current then gradually turns off and the ChR

'recovers', whereupon the whole process can be repeated. But the precise mechanism for how light opens the channel gate and how the gate closes is not known. On page 369 of this issue, Kato *et al.*² report a high-resolution X-ray crystal structure of a genetically engineered ChR, and use it to propose an explanation for how the isomerization of its chromophore causes pore opening.

Although the behaviour of motile algae has been studied for decades, it wasn't until 2002 that an eyespot pigment of the alga *Chlamydomonas reinhardtii* was identified³ as the light-activated protein channelrhodopsin 1. Three years later, ChRs were expressed in mammalian neurons and used to facilitate the light-induced stimulation of the cells' activity⁴. Subsequent bioengineering of ChRs, enabling optical control of cells on the millisecond timescale, together with the development of systems for delivering genes to specific cell types, boosted the rapidly growing field of optogenetics⁵. Since then, the use of this technology has grown exponentially, with no signs of its popularity waning.

Microbial opsins — the family of light-activated proteins that includes channelrhodopsin — have been the most commonly used protein tools for optogenetics⁶. The light sensitivity and spectral absorption of opsins are due to the fact that the proteins are covalently bound to their chromophore (all-*trans*-retinal, a derivative of vitamin A). These proteins share a common structural plan, which includes seven transmembrane helices and a characteristic bond (known as a Schiff base) that connects retinal to a lysine amino-acid residue in helix 7.

The first microbial opsins to be identified were bacteriorhodopsin and halorhodopsin, both of which were found in halobacteria. Bacteriorhodopsin (BR) uses light energy to pump protons out of cells, whereas halorhodopsin pumps chloride ions in the opposite direction. In halobacteria, a complex of a sensory rhodopsin and a transducer protein mediates phototaxis (the microbe's movement in response to light). But in microalgae, ChR performs this task without a transducer, opening its pore in response to light to generate an ion current. It can do this because the light-sensitive chromophore and the channel reside on the same polypeptide chain.

What puts these molecular channels and pumps in the top drawer of the optogenetics toolbox for neuroscientists is the fact that they allow light to be used as a fairly innocuous method to change the ion gradient across the membrane of a neuron, thereby enabling cell depolarization (neuronal activation) or cell hyperpolarization (neuronal silencing). The latest optogenetic gadgets actually contain two microbial opsins linked in tandem, a system that allows greater control of ion flow compared with previously used individual opsins⁷.

Although a wealth of structural and biophysical studies have improved our understanding of the pump processes for BR and halorhodopsin, relatively little is known about the gating process of ChR. What is known is that the ChR process, like those of BR and halorhodopsin, is cyclic, with each cycle lasting tens of milliseconds and involving several intermediates. Experiments that introduced targeted mutations into ChRs, and analysed the proteins' electrophysiological and spectroscopic properties, have also yielded modified ChRs that show altered ion preferences, spectral properties and pore-opening and -closing kinetics.

Kato *et al.*² now report that the structure of ChR (Fig. 1), although similar to that of BR in some respects, also brings a few surprises. These unexpected features might explain the properties of some of the commonly used engineered ChR mutants. When the authors superimposed the structure of their ChR on that of BR, they found that the transmembrane domain and position of retinal are similar. But unlike BR, which assembles in trimers in the membrane of halobacteria, ChR forms a dimer in which the two subunits are in

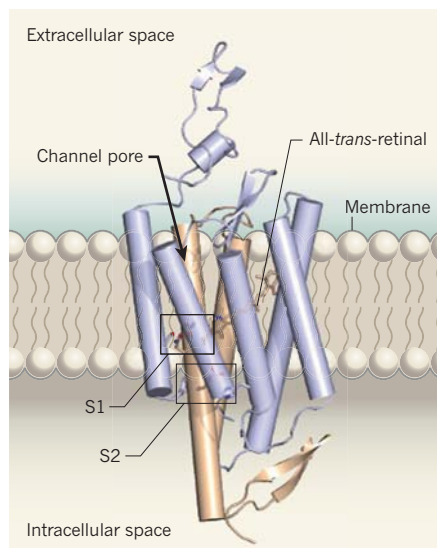


Figure 1 | Structure of a closed light-gated cation channel. Channelrhodopsins (ChRs) are proteins that form channels in microbial cell membranes. The channels form from seven transmembrane helices (shown as cylinders) and open in response to light, allowing cations to pass through the membrane. Their light sensitivity is caused by a molecule, all-*trans*-retinal, that is covalently attached to the protein. Kato *et al.*² report the X-ray crystal structure of a chimaeric ChR constructed from two other ChRs, ChR1 and ChR2; the purple parts of the protein are from ChR1 and the brown parts are from ChR2. The chimaeric ChR forms dimeric structures, but only one ChR is depicted. The authors find that a negatively charged pore sits between helices 1, 2, 3 and 7 and is interrupted by two trios of amino acids, which form gates S1 and S2. They propose that light-induced isomerization of the retinal causes the gates to open, extending the pore to the cytoplasm.

close contact, in agreement with a previously proposed structure⁸ obtained using electron crystallography.

Another difference between BR and ChR is that ChR has extended amino-terminal and carboxy-terminal domains. The N-terminal extension contains three cysteine amino-acid residues, which form covalent disulphide bonds with their counterparts in a second ChR molecule, enabling dimerization. The C-terminal extension forms a β -sheet at the end of the long helix 7, which protrudes into the intracellular space. For their study, Kato *et al.* crystallized a truncated version of ChR, which — like ChRs used as optogenetic tools — consists only of the transmembrane part and lacks more than half of the naturally occurring protein. So, the β -sheet observed by the authors may be a part of the large, mostly missing C-terminal domain, which is thought to be involved in subcellular localization and tethering of the ChR to the algal eyespot.

Perhaps the most notable difference between ChR and BR is that the extracellular ends of helices 1 and 2 in ChR are tilted outward by about 3–4 Ångströms with respect to the

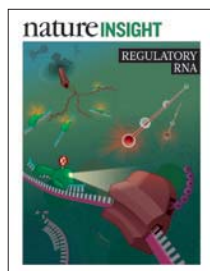
analogous helices in BR. Together with helices 3 and 7, this creates a pore extending halfway through the protein. The authors observe that the inside surface of the pore contains many negatively charged amino-acid residues. Most of these are glutamic acid residues from the extracellular part of helix 2, suggesting that this helix is mainly responsible for defining the pore's conductance and ion selectivity. The negatively charged pore elongates into a slightly positively charged vestibule in the extracellular space.

The pore created by helices 1, 2, 3 and 7 ends in the middle of the protein, where retinal resides. On its intracellular side, the pore is constricted by two gates, each consisting of three residues from different helices. The three residues of the innermost gate form a hydrogen-bonding network next to retinal's attachment site, suggesting that helix movements caused by retinal's isomerization might break the network and open the gate. The movement of helices 1 and 2 might also open a second gate further towards the intracellular side, along the putative cation channel. Although Kato *et al.*² argue quite convincingly that the cation-conductance pore comprises helices from a single ChR molecule, an alternative hypothesis is that the ChR dimer assembles to form the pore using elements from each of the two ChRs⁸. Such an arrangement would be reminiscent of the situation reported⁹ for two-pore-domain potassium channels.

With this first report² of a high-resolution crystal structure of an engineered ChR, are we at the dawn of the age of 'structural optogenetics'? Structure-based design, combined with the discovery of other useful microbial ChRs, might produce optogenetic tools that have highly specific properties tailored for the study of individual cell types and signalling processes. And although it is too early to say "take photons, not drugs", the potential for optogenetics to revolutionize neuroscience and neurology is now even more in the spotlight. ■

Oliver P. Ernst is in the Departments of Biochemistry and Molecular Genetics, University of Toronto, Toronto, Ontario M5S 1A8, Canada. **Thomas P. Sakmar** is in the Laboratory of Molecular Biology and Biochemistry, The Rockefeller University, New York, New York 10065, USA. e-mails: oliver.ernst@utoronto.ca; sakmar@rockefeller.edu

1. Miesenböck, G. *Annu. Rev. Cell Dev. Biol.* **27**, 731–758 (2011).
2. Kato, H. E. *et al. Nature* **482**, 369–374 (2012).
3. Nagel, G. *et al. Science* **296**, 2395–2398 (2002).
4. Boyden, E. S., Zhang, F., Bamberg, E., Nagel, G. & Deisseroth, K. *Nature Neurosci.* **8**, 1263–1268 (2005).
5. Deisseroth, K. *Nature Meth.* **8**, 26–29 (2011).
6. Zhang, F. *et al. Cell* **147**, 1446–1457 (2011).
7. Kleinlogel, S. *et al. Nature Meth.* **8**, 1083–1088 (2011).
8. Müller, M., Bamann, C., Bamberg, E. & Kühlbrandt, W. *J. Mol. Biol.* **414**, 86–95 (2011).
9. Poulsen, H. & Nissen, P. *Science* **335**, 416–417 (2012).



Cover illustration
by Nik Spencer

Editor, *Nature*
Philip Campbell

Publishing
Nick Campbell

Insights Editor
Ursula Weiss

Production Editor
Nicola Bailey

Art Editor
Nik Spencer

Sponsorship
Gerard Preston

Production
Emilia Orviss

Marketing
Elena Woodstock
Hannah Phipps

Editorial Assistant
Hazel Mayhew

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel: +44 (0) 20 7833 4000
e: nature@nature.com



nature publishing group

Although proponents of RNA might beg to differ, in the hierarchy of popular interest, DNA has historically held more sway. Being able to decipher genomes was seen as a milestone on the way to understanding life itself. What genome-wide RNA sequencing studies have revealed, however, is the unexpected complexity of RNA species encoded by DNA, most of which do not code for a protein. We now appreciate that such non-coding RNAs exert important regulatory controls on many biological processes.

The reviews in this Insight illustrate some of these principles. RNA is synthesized as a single-stranded molecule, but it is able to base-pair with itself, other RNA molecules or DNA. Hashim Al-Hashimi and colleagues discuss how secondary and tertiary structures of RNA are influenced by external cues to elicit a specific functional output. The cell exploits this dynamism to regulate processes such as transcription, post-transcriptional processing and translation. Jennifer Doudna and colleagues review a microbial adaptive immune system, CRISPR (clustered regularly interspaced short palindromic repeat). This system incorporates small pieces of invading viral or plasmid sequences into the bacterial genome as CRISPR loci; when future invasions occur, the expressed CRISPR RNAs recognize the foreign nucleic acids and mediate their degradation. The physiological function of many long non-coding RNAs remains undetermined, but Mitchell Guttman and John Rinn propose a model in which these molecules act in a modular fashion to bind different proteins or hybridize to various DNAs or RNAs; this modularity expands the scope of a single RNA's function. Finally, Amaia Lujambio and Scott Lowe highlight the role of another class of much shorter, non-coding RNA — microRNAs — in cancer development and suppression, and as a target for therapeutic intervention.

We hope these reviews provide a flavour of how the inherent properties of RNA make it a robust species to regulate cellular processes.

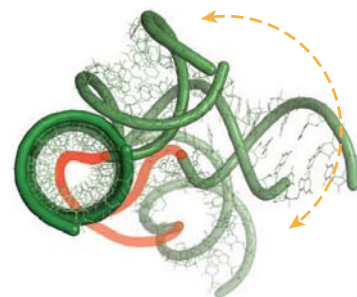
**Angela K. Eggleston, Alex Eccleston,
Barbara Marte & Claudia Lupp**
Senior Editors

CONTENTS

REVIEWS

322 Functional complexity and regulation through RNA dynamics

Elizabeth A. Dethoff, Jeetender Chugh, Anthony M. Mustoe & Hashim M. Al-Hashimi



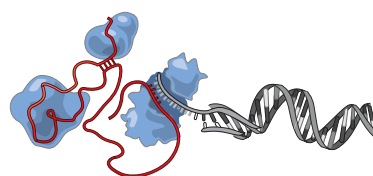
331 RNA-guided genetic silencing systems in bacteria and archaea

Blake Wiedenheft, Samuel H. Sternberg & Jennifer A. Doudna



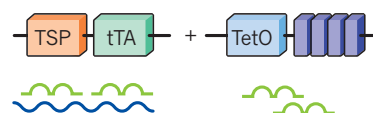
339 Modular regulatory principles of large non-coding RNAs

Mitchell Guttman & John L. Rinn



347 The microcosmos of cancer

Amaia Lujambio & Scott W. Lowe



Functional complexity and regulation through RNA dynamics

Elizabeth A. Dethoff¹, Jeetender Chugh¹, Anthony M. Mustoe¹ & Hashim M. Al-Hashimi¹

Changes to the conformation of coding and non-coding RNAs form the basis of elements of genetic regulation and provide an important source of complexity, which drives many of the fundamental processes of life. Although the structure of RNA is highly flexible, the underlying dynamics of RNA are robust and are limited to transitions between the few conformations that preserve favourable base-pairing and stacking interactions. The mechanisms by which cellular processes harness the intrinsic dynamic behaviour of RNA and use it within functionally productive pathways are complex. The versatile functions and ease by which it is integrated into a wide variety of genetic circuits and biochemical pathways suggests there is a general and fundamental role for RNA dynamics in cellular processes.

Analysis of the first X-ray structure of the protein myoglobin¹ prompted researchers to ask the question: how do ligands reach the deeply buried haem group centre? This simple, but powerful, observation has inspired decades of investigation into the dynamic behaviour of proteins, so that we now know protein structures are in constant motion, and that these fluctuations in structure are crucial to, and sometimes drive, function. Early X-ray structures of RNA contained indications of the importance of conformational dynamics: large changes in the helical arms of transfer RNA were observed on the binding of tRNA synthetase², and changes in the conformation of ribozymes needed to be invoked to envision catalytically active states^{3–5}. However, no one could have anticipated the existence of new genetic circuits that are based on RNA conformational switches, or that the ‘acrobatic’ nature of a biopolymer that consists of only four chemically similar nucleotides would be at the centre of a complex macromolecular structure such as the ribosome.

The dynamic changes that occur in the structure of RNA serve an ever-increasing range of functionality that generally follows a common two-step process (see Supplementary Information for more reviews on RNA dynamics). The process involves a cellular signal that triggers RNA dynamics, which are then transduced into a specific biological output. This Review provides a critical account of RNA dynamics as a regulatory mechanism and source of functional complexity. We review the known dynamic properties of RNA structure and emphasize the unique properties that allow large changes in structure to take place in a biologically specific and robust manner. We then examine the wide range of cellular inputs used to interface with RNA dynamics and the various mechanisms that are used to guide the dynamics to achieve a broad spectrum of functional outputs.

RNA free-energy landscape

It is important to distinguish between the two types of RNA dynamics: ‘equilibrium fluctuations’ and ‘conformational transitions’. Equilibrium fluctuations are related to the thermal activated motions that occur in RNA. Conformational transitions arise when cellular cues, such as an increase in the concentration of a metabolite, create a non-equilibrium state that then relaxes back to equilibrium. This Review is focused principally on conformational transitions because of their dominant role in regulatory mechanisms; however, the two motions are intricately related, as highlighted by studies of RNA and protein

dynamics^{6,7}. This, and other aspects of RNA dynamic behaviour that are relevant to function, is best understood by looking at the free-energy landscape of RNA^{8,9}.

The free-energy landscape specifies the free energy of every possible RNA conformation (Fig. 1a). Equilibrium fluctuations correspond to the spontaneous jumps that occur between various conformers along the free-energy landscape. The population of a given conformer depends on its free energy, whereas the transition rate between conformers depends on the free-energy barrier of separation (Fig. 1a). Conformational transitions arise when cellular cues perturb the free-energy landscape, which leads to a redistribution of conformational states (Fig. 1a). The RNA free-energy landscape is punctuated by deep local minima, or conformational wells, in which conformations within a well are highly similar and conformations from different wells are structurally distinct. These are the conformations that are significantly sampled by equilibrium motions and are stabilized by cellular cues to effect conformational transitions^{10–12} (Fig. 1a). For example, the degeneracy of base-pairing and stacking interactions, together with the high stability of RNA duplexes, results in deep local minima that correspond to different but energetically equal secondary structures that are separated by large kinetic barriers¹³ (Fig. 1b). As few as two secondary structures can dominate the RNA dynamic landscape because the loss of energy that accompanies the disruption of just one base-pair can markedly destabilize alternative conformations. In addition, RNA of a given secondary structure can undergo more facile dynamic excursions in tertiary structure, which involve smaller energetic barriers. These dynamics are commonly dominated by large changes in the relative orientation of helical domains, which carry motifs involved in tertiary contacts, and occur around flexible pivot points consisting of bulges, internal loops and higher-order junctions (Fig. 1c).

Although these excursions can lead to very large changes in tertiary structure, they are limited to a narrow set of conformations. For example, calculating the set of conformations that are accessible to two helices that are connected by a three-residue bulge reveals that the interhelical bend angle, when combined with interhelical twisting, can range from 0° to 180°. Despite this large accessible range, the connectivity constraints that are imposed by the bulge junction and the steric forces that act on the two helices direct changes in the interhelical orientations along a highly directional pathway and therefore restrict the conformational space to less than 20% of what

¹Department of Chemistry and Biophysics, The University of Michigan, 930 North University Avenue, Ann Arbor, Michigan 48109-1055, USA.

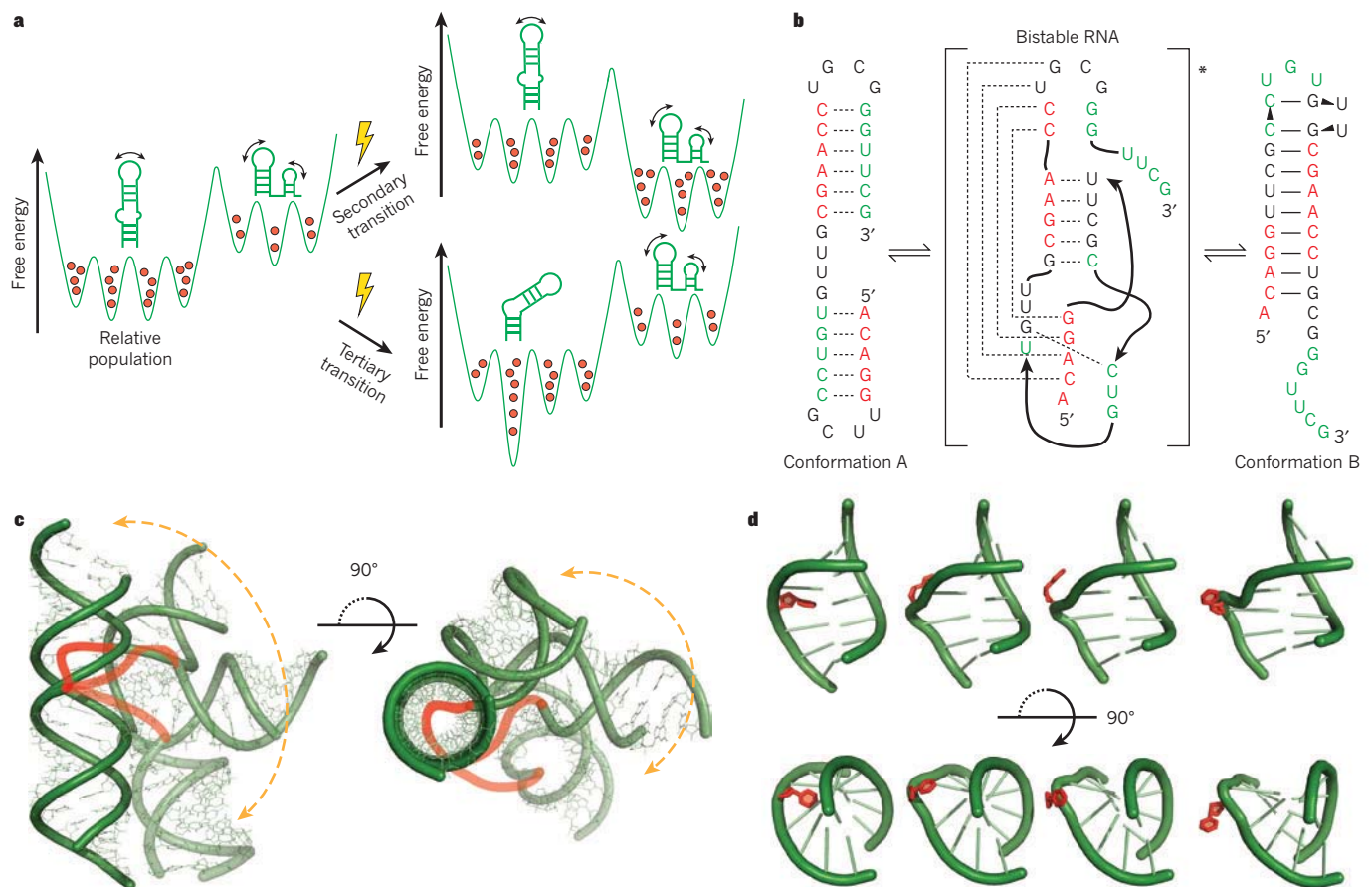


Figure 1 | Shape and form of RNA dynamics. **a**, The secondary and tertiary RNA conformations of different low-lying energy states are shown above the RNA free-energy landscape (green line). The relative populations of each conformation are shown within the landscape (red balls). Cellular effectors (bolts) can modify the energy landscape to favour an alternative secondary structure (top), or preferentially stabilize an alternate tertiary conformation (bottom). **b**, Exchange between alternative, isoenergetic secondary structures (A and B) that are separated by large energetic barriers owing to disruption of base pairs in the transition state¹³. **c**, The accessible range of interhelical

conformations for an RNA two-way junction consisting of a trinucleotide bulge, with the possible paths of the bulge, which were excluded during the modelling, illustrated in red^{14,15}. The allowed range of conformations is restricted towards a specific and directed conformational pathway by steric and stereochemical forces. The structure is rotated 90° to illustrate the bending (left) and twisting (right) motion. **d**, Flipping out of a residue (red) participating in a non-canonical base pair within an RNA internal loop is illustrated, progressing from an intrahelical stacked to an extrahelical unstacked conformation. The motion occurs without perturbing flanking Watson–Crick pairs (green).

is theoretically possible^{10,14–16} (Fig. 1c). In addition, owing to the high stability of duplexes, residues participating in non-canonical base pairs can loop-out from intra- to extrahelical conformations without significantly disturbing the structure of the flanking helices^{17,18} (Fig. 1d). Precise control over these dynamics is encoded within the sequence, and small sequence variations can greatly alter the relative populations of different RNA structures and their rates of interconversion^{11,19}. For example, distinct interhelical orientations can be sampled by changing the length and asymmetry of junctions^{10,14,15}, and the tendency of residues to loop-out can be modulated on the basis of sequence-specific stacking interactions^{20,21} (see Supplementary Information for links to movies and animations of experimentally determined RNA dynamics).

These features can help to explain the three remarkable aspects of RNA conformational transitions that are of fundamental importance for regulatory functions. First, the landscape is hierarchical due to the height of the energy barriers that separate alternative secondary structures. Changes in tertiary contacts rarely involve changes in the secondary structure and the two types of conformational changes can be used to serve different functions. Throughout this review, we will use ‘secondary’ and ‘tertiary’ conformational changes to distinguish between these two types of dynamics. Second, the limited landscape of energetically favourable conformations allows RNA to undergo very large changes in structure, but to be directed towards a very specific set of conformations from a vast number of possibilities. Third,

there is increasing evidence that RNA dynamics are determined by the underlying RNA free-energy landscape, and to lesser extent by cellular cues^{7,22,23}. Thus, conformational transitions can be considered perturbations that guide pre-existing equilibrium fluctuations towards specific functionally productive pathways. In this way, even an imperfect force or cellular signal will drive changes in the RNA structure along a predetermined pathway, which makes the transitions highly robust.

Triggers of RNA conformational transitions

RNA dynamics can be triggered by a remarkably diverse set of molecular effectors and environmental cues through several different mechanisms. This provides many different points of entry for integrating RNA conformational transitions into biological circuits and biochemical pathways.

Specific protein binders

The most common effectors are proteins that bind their target RNA specifically through well-defined structural features, thereby stabilizing one or a subset of conformations from the pre-existing energy landscape. For example, the mitochondrial tyrosyl-tRNA synthetase CYT-18 from *Neurospora crassa* binds specifically to group-I introns (a class of large self-splicing ribozymes that catalyse their own excision from messenger RNA, tRNA and ribosomal RNA precursors) and stabilizes the conformation required for catalytic activity²⁴. Protein binding often leads to large changes in the overall orientation of RNA helices around

junctions such as bulges²⁵, three-way junctions²⁶ and other motifs such as the K-turn²⁷. For example, the spliceosomal U4 small nuclear RNA (snRNA) undergoes a sharp transition in the interhelical bend angle, from approximately 69° to 25°, around a K-turn motif, when it binds to its cognate protein target²⁸ (Fig. 2a). These changes in interhelical conformation are driven in part by nonspecific electrostatic interactions between basic amino acids and the high negative-charge density that builds up at interhelical junctions, and are often observed as equilibrium dynamics in the absence of an effector^{29–31}. For example, unbound HIV-1 transactivating response RNA (TAR) dynamically samples the many different interhelical orientations that are observed when it is bound to seven distinct ligands, including peptide mimics of its cognate protein, Tat³¹ (Fig. 2b).

In an increasing number of cases, protein binding does not involve the stabilization of a specific minimum of the RNA free-energy landscape. Instead, binding selectively lowers the surrounding energy barriers to accentuate or alter the equilibrium dynamics of the RNA. For example, binding of the U1A protein to its cognate RNA target does not cause the pre-existing equilibrium interhelical motions to stop, but rather induces mobility in regions of the RNA that are in direct contact with the protein³². The CBP2 protein from yeast mitochondria binds specifically to the bI5 group-I intron and activates large-scale RNA equilibrium motions³³. Even simple small-molecule ligands lead

to the reorganization of the TAR RNA equilibrium dynamics³⁴. These observations highlight the importance of embracing a broader view of trigger factors as elements that perturb the entire energy landscape and guide RNA dynamics rather than simply stabilize a single conformation from a dynamic range.

RNA chaperones and helicases

As is often the case in RNAs that possess alternative secondary structures, the large energy barriers associated with base-pair melting can limit the dynamics between RNA conformational wells. In this way, the RNA can become kinetically trapped in a metastable, non-equilibrium conformation. In response to this, a variety of proteins have evolved that possess the RNA 'chaperone' activity needed to efficiently drive RNA secondary structural-transitions over the large energy barriers^{35,36}. For example, the HIV nucleocapsid protein uses non-specific interactions between the RNA and protein to destabilize the RNA helices³⁷. This lowers the energetic barrier to conformational exchange, accelerating relaxation to equilibrium and allowing metastable RNAs to convert to conformations that are more thermodynamically favourable.

Other RNA chaperones, such as RNA helicases, help RNA traverse high energy barriers by unwinding helices and disrupting RNA structure, as well as promoting the formation of new RNA duplexes to accelerate conformational transitions in RNAs and ribonucleoprotein (RNP)

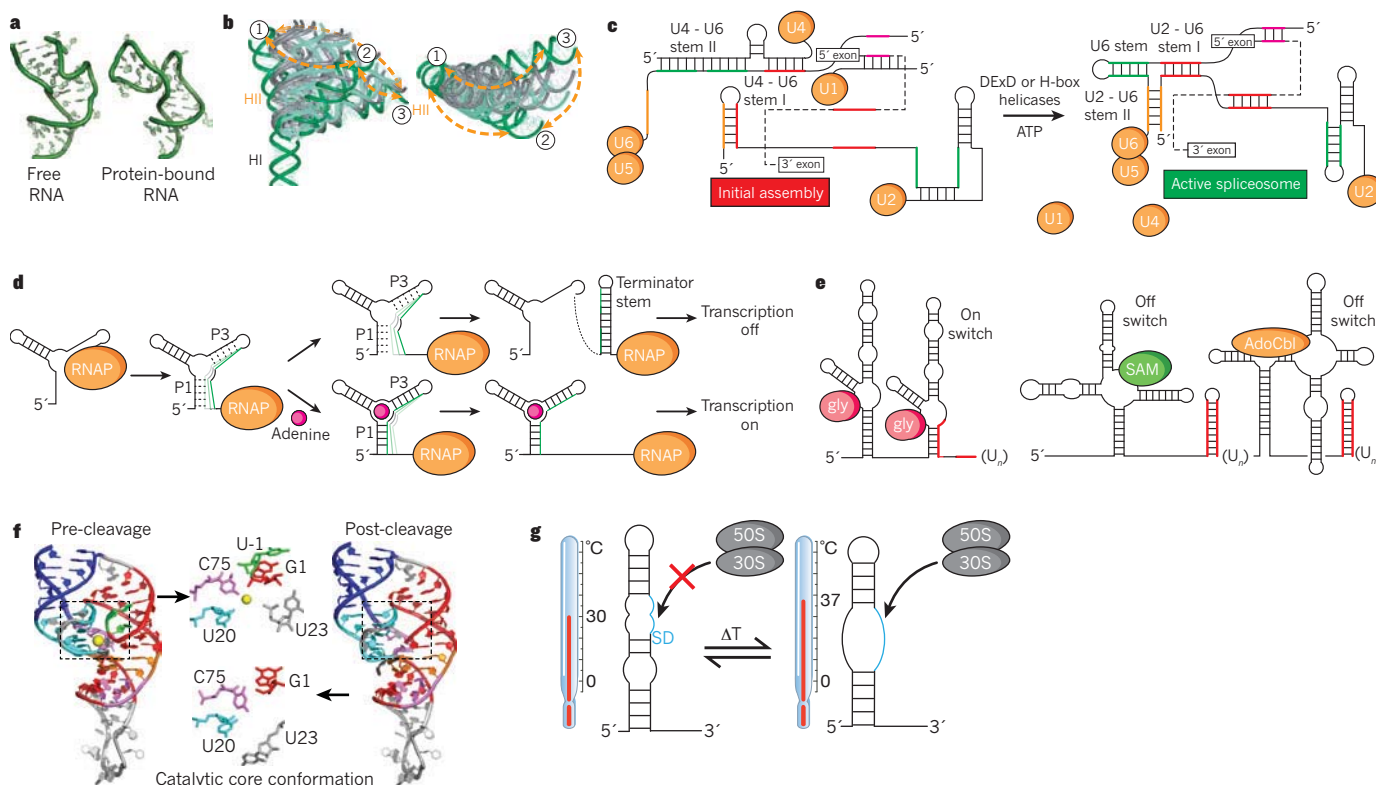


Figure 2 | Triggering RNA conformational transitions. **a**, Conformational changes in the spliceosomal U4 snRNA K-turn motif (Protein Data Bank (PDB) ID 2KR8) triggered when it binds to a complex of the human protein PRP31 and the 15.5K protein (PDB ID 2OZB)²⁸. **b**, Similarity between the TAR RNA interhelical conformations that are triggered by binding to small molecules, Tat peptide derivatives and divalent ions (grey helices) and those that are sampled by equilibrium dynamics (green helices labelled as 1, 2 and 3) in the unbound state shown as a horizontal and vertical view. The path of helix II (HII) as it moves from one unbound, equilibrium conformer to the next is shown by the orange arrows. HI, helix I. Figure modified, with permission, from ref. 31. **c**, RNA conformation transitions during spliceosome assembly on pre-mRNA (dashed line) in the presence of DExD or H-box helicases and ATP. Sections are colour-coded to indicate base-pairing after helicase action. **d**, The RNA structure is modulated by steering of the co-transcriptional folding pathway. The adenine transcription-terminating riboswitch is a typical example. The progression of co-transcriptional

folding with and without the ligand (adenine) is shown. Adenine binds to the aptamer domain and stabilizes the structure, allowing transcription to be turned on. RNAP, RNA polymerase. **e**, Two examples of tandem riboswitch architectures. Left, cooperative binding of glycine by the glycine riboswitch using tandem aptamer domains and one expression platform. Right, tandem SAM and AdoCbl riboswitches in which either of the two ligands triggers the conformational switch and yields an output of gene repression. Sequences that can form transcription terminator stems are shown in red. **f**, Conformations of HDV ribozyme pre-cleavage (PDB ID 1VC7)⁵³ and post-cleavage (PDB ID 1DRZ)⁵² states. Enlarged details of the catalytic core (dashed box) of the two structures are shown, with the bound substrate (green) and the magnesium ion (yellow) present only in the pre-cleavage state. **g**, Melting of the secondary structure around the ribosome-binding site of virulence genes in the pathogen is triggered by an increase in temperature that makes the Shine-Dalgarno sequence (SD, blue) available for ribosome binding and translation initiation.

complexes³⁸. These chaperone proteins are important for remodelling the structure of RNA and RNP complexes because they can anneal or unwind RNA strands depending on the environmental cues³⁹. For example, helicases are essential in the assembly of the spliceosome, which is a complex RNP that consists of five RNAs and multiple proteins that catalyses excision of introns from a nuclear pre-mRNA^{40,41}. Assembly proceeds through a series of transitions that involve the melting and annealing of RNA duplexes that are catalysed by DExD/H-box ATPase helicases (Fig. 2c). For example, the U4 RNA escorts the U4–U6–U5 triple small nuclear RNP complex (tri-snRNP) to the pre-mRNA, but is subsequently released by the DExD/H-box helicase Brr2, which catalyses the melting of the two stems within U4 and U6. This frees the U6 stem to base-pair with U2 snRNA and leads to a new RNA structure that is required for the first transesterification reaction⁴² (Fig. 2c). In addition, DExD/H-box proteins are involved in the release of mRNA produced in pre-mRNA splicing reactions. For example, the DEAH-box splicing factor Prp22 is deposited on spliced mRNA downstream of the exon–exon junction and catalyses the disruption of contacts between mRNA and U5 snRNP, thereby releasing the spliced mRNA from the U5–U6–U2 spliceosomal assembly⁴³. In another example of the variety of functions of RNA chaperones, the DExD/H-box protein CYT-19 unfolds native and misfolded conformations of a group-I catalytic RNA in an ATP-dependent process. A large free-energy gap between the native and misfolded conformers directs CYT-19 to unfold misfolded conformers more frequently than native conformers. In the process, CYT-19 redistributes the two conformation populations, which allows native RNA to populate a wider range of conformations than would otherwise be possible⁴⁴.

Metabolites and physiochemical conditions

Another ingenious strategy is used to modulate RNA structure in response to a wide range of metabolite-based effectors, including small molecules (such as amino-acids, coenzymes and nucleotides^{23,45}) and changes in physiochemical conditions (such as magnesium ion concentration⁴⁶ and pH⁴⁷). It would be difficult, if not impossible, for these smaller effectors and cellular cues to possess the chaperone activity needed to efficiently drive secondary structural transitions over the associated large energy barriers. Instead, this strategy operates on the initial RNA-folding process itself, intervening while the energy barriers are still low. Specifically, these effectors and cues act by directing the RNA to different folding pathways during RNA co-transcriptional folding. This process is made possible by the unidirectional and comparatively slow rate with which RNA is transcribed from the 5' to the 3' direction relative to RNA folding and effector binding. Each pathway favours one of two distinct secondary structures, where each secondary structure is associated with an alternative biological outcome (Fig. 2d). This trigger mechanism is implicated in a growing list of other RNA switches, although it has been best described for metabolite-sensing riboswitches^{23,45}.

Riboswitches are RNA-based genetic elements typically embedded in the 5' untranslated region of bacterial genes that regulate expression of metabolite genes in response to changes in cellular metabolite concentration^{23,45}. In a prototypical metabolite riboswitch, a metabolite, such as adenine, binds to the aptamer domain with exceptional affinity and selectivity. This stabilizes an otherwise shallow energy well, which induces a redistribution of the aptamer conformational states towards one state that, in most riboswitches, sequesters an RNA element into a helix of the aptamer domain⁴⁸ (Fig. 2d). In turn, the unavailability of the RNA element changes the folding pathway of a downstream decision-making expression platform, directing it towards structures that turn off (and in some cases, turn on) gene expression, either by forming a transcription-terminating helix (Fig. 2d) or by sequestering the Shine-Dalgarno sequence (a ribosome binding site located eight base pairs upstream of the start codon in mRNA), thereby inhibiting translation. This system also keeps the number of spontaneous conformational transitions, or premature switching in the absence of ligands, to a minimum because very large energy barriers separate the two alternative

secondary structural forms of the expression platform.

More complex functionality can be achieved by coupling multiple riboswitches together. For example, the glycine riboswitch uses two aptamer domains in tandem to cooperatively bind glycine, thereby increasing responsiveness to changes in ligand concentrations⁴⁹ (Fig. 2e). The tandem arrangement of two entire riboswitches that respond to two distinct ligands allows the construction of more sophisticated genetic circuits such as two-input Boolean NOR logic gates, in which either of the two ligands can trigger the conformational switch and yield an output of gene repression⁵⁰ (Fig. 2e). In another example, the c-di-GMP-sensing riboswitch and a GTP-dependent self-splicing group-I ribozyme in the 5' untranslated region of a putative *Clostridium difficile* virulence gene work in tandem to regulate translation⁵¹. In the presence of c-di-GMP and GTP, the riboswitch and ribozyme form a structure that stabilizes a 5' splice site, and the ribozyme self-splices to yield an RNA transcript with a perfect ribosome-binding site located upstream of the start codon. Conversely, in the presence of GTP alone, alternative base pairing between the riboswitch and ribozyme occurs to form a structure that promotes splicing at an alternative site, which results in a splicing product without a ribosome-binding site, and thus downregulates translation. This RNA arrangement represents the first natural example of an allosteric ribozyme.

Chemical reactions

Chemical reactions, such as cleavage of the RNA phosphodiester backbone, can also reshape the underlying RNA energy landscape so that a state that was previously in equilibrium becomes a non-equilibrium state, which triggers changes in RNA secondary and tertiary structure. For example, X-ray analysis of the structures of precursor and product states of the hepatitis delta virus (HDV) ribozyme, which catalyses site-specific self-cleavage of the viral RNA phosphodiester backbone, reveal changes in the local arrangement of catalytic groups, as well as the ejection of a catalytically important magnesium ion⁵². These conformational changes may help to accelerate product release^{53,54} (Fig. 2f). Another example is seen in the secondary structural switch triggered by cleavage of the 3' end of the pre-18S rRNA during eukaryotic ribosome maturation, which is used to enforce a sequential order to the maturation process⁵⁵.

Thermal and mechanical triggers

Other energy-dependent processes can induce the complete 'melting' of RNA helices. RNA thermosensors alter expression of genes during heat-shock response and pathogenic invasion in response to increases in temperature⁵⁶ (Fig. 2g). For example, when *Listeria monocytogenes* invades an animal host, the pathogen enters a warmer environment, which activates a thermosensor located at the 5' untranslated region of the *prfA* mRNA⁵⁷. The higher host temperature causes a shift in the energy landscape from one that favours the formation of the thermosensor hairpin to one that favours the melted, single-stranded conformation. This melting transition exposes ribosome-binding sites, which are required for translation. Mechanical triggers can also induce the unfolding of RNA hairpins. One example is translation-induced unfolding of mRNA hairpins, which is thought to slow the rate of ribosome elongation to allow the folding of autonomous-folding proteins and protein domains⁵⁸.

Functions of secondary structural transitions

Secondary structural transitions are widely used in gene regulation as binary switches that are activated by cellular cues. The switch can be transduced into a range of outputs by sequestering or exposing key RNA regulatory elements.

Transcription

Many RNA switches regulate gene expression at the transcriptional level by producing transcription-terminating helices. In addition to metabolite-sensing riboswitches, other RNA switches use the same strategy to regulate gene expression in response to more complex molecules^{23,45}. For example, in the T-box mechanism,

non-aminoacylated or uncharged tRNA can activate transcription of the cognate gene that encodes its aminoacyl-tRNA synthetase. The interaction between the acceptor end of an uncharged tRNA and residues in the antiterminator bulge in the 5' untranslated region of the mRNA (Fig. 3a) promotes formation of an antiterminator helix during co-transcriptional folding that allows transcription to continue. However, the acceptor end of aminoacylated or charged tRNA cannot interact with the antiterminator helix residues, which results in formation of the more stable terminator stem that aborts synthetase gene transcription⁵⁹ (Fig. 3a). Only a few proteins have been identified that modulate transcription by influencing folding of transcription-terminating helices. One example is the tryptophan-activated RNA-binding attenuation protein (TRAP), which binds *trp* mRNA to regulate gene expression at both the transcriptional and translational level by several processes (for example, promoting the formation of a terminator hairpin that terminates transcription⁶⁰).

Translation

There is an increasing list of protein- and RNA-triggered⁶¹ RNA switches that regulate translation by sequestering or exposing ribosome-binding sites or by affecting the structure of ribosomal RNA, and therefore blocking translation. For example, a protein-dependent RNA switch has recently been identified in the 3' untranslated region

of *VEGFA* mRNA in myeloid cells that regulates translation of *VEGFA* in response to proteins associated with two disparate stress stimuli (Fig. 3b). The interferon- γ (IFN- γ)-activated inhibitor of translation (GAIT)-complex binds a structural GAIT element within a family of inflammatory mRNAs and silences their translation by promoting the formation of a translational-silencing (TS) conformer⁶². During oxidative stress, the heterogeneous nuclear ribonucleoprotein L (hnRNP L) overrides GAIT silencing by triggering a secondary structural RNA switch to a translation-permissive (TP) conformer, in which the GAIT element is occluded. The RNA alternates between two mutually-exclusive conformers in response to the binding of the GAIT complex or hnRNP L, thereby functioning as an AND NOT Boolean logic-gate switch in which the presence of one protein, but not the other, yields an output of gene repression (Fig. 3b).

Post-transcriptional processing

An increasing number of RNA switches are involved in regulating post-transcriptional processing; for example, splicing, gene silencing by microRNA (miRNA) and RNA editing. Although the detailed mechanics of many of these systems are still unknown, in all cases the RNA switch exposes, occludes or modulates the structure of the processing sites to regulate post-transcriptional processes. For example, one of the thiamine pyrophosphate riboswitches discovered in eukaryotes

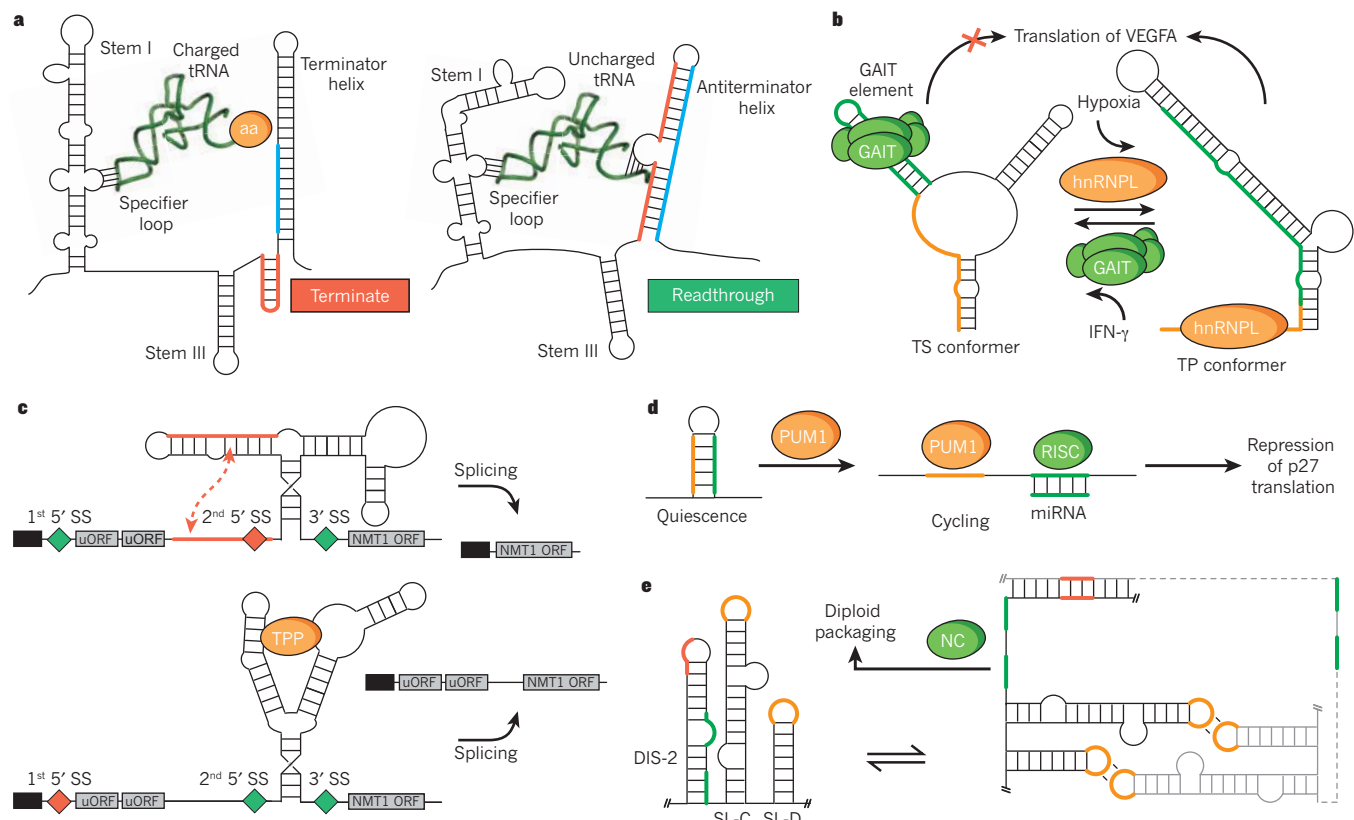


Figure 3 | Functional outputs of secondary structural changes. **a**, Transcriptional activation of the aminoacyl-tRNA synthetase gene by uncharged tRNA (no aminoacylation (aa)). Binding of uncharged tRNA induces formation of an antiterminator helix during co-transcriptional folding⁵⁹. **b**, Translation control of *VEGFA* expression through a dual protein-dependent RNA secondary structural switch that responds to interferon- γ (IFN- γ) by binding the IFN- γ -activated inhibitor of translation (GAIT) complex (green) to form a translational-silencing (TS) conformer (on the left) and to hypoxic stress that results in hnRNP L binding and causes a switch to a translation-permissive (TP) conformer (on the right). **c**, Thiamine pyrophosphate (TPP) riboswitch-regulated alternative splicing and gene expression of NMT1. In the absence of TPP, the aptamer domain base-pairs (red dotted line) to the sequence surrounding a proximal 5' splice site (SS,

shown as coloured diamonds: green, activation or red, repression) to block it from the SS machinery. Instead, a distal SS is selected. On binding TPP, the aptamer domain undergoes a conformational change to expose the second proximal 5' SS. The resultant spliced mRNA contains decoy upstream open reading frames (uORFs), thus reducing expression of the NMT1 ORF. **d**, Pumilio protein-mediated mRNA secondary structural switch controls accessibility of microRNA-binding sites and regulates expression of p27 protein. Binding of PUM1 induces a conformational change to expose the miR-211 and miR-222 binding site to allow p27 silencing. RISC, RNA-induced silencing complex. **e**, Secondary structural switch couples dimerization and diploid genome packaging of the Moloney murine leukaemia virus. Dimerization leads to a coupled frame-shift that exposes nucleocapsid protein binding sites (green) required for genome packaging. NC, nucleocapsid.

regulates alternative splicing⁶³ (Fig. 3c). Here, changes in the secondary structure sequester or expose splice sites (Fig. 3c). An RNA switch has recently been identified in the 3' untranslated region of p27 mRNA that simultaneously sequesters both an miRNA target site from cleavage by the RNA-induced silencing complex (RISC) and a Pumilio-recognition element (PRE), which binds a Pumilio RNA-binding protein (PUM1)⁶⁴. Binding of PUM1 to the PRE region triggers a secondary structural switch that exposes the miRNA target site, leading to miRNA silencing (Fig. 3d). In another example, HDV genotype III editing levels are determined by a pre-existing equilibrium between two secondary structures of the antigenome RNA, involving a kinetically trapped conformation and a thermodynamically more favourable state⁶⁵. These initial discoveries suggest RNA switches have a range of functions in post-transcriptional processing.

Viral replication

RNA genomes of retroviruses take advantage of RNA secondary structural switches to transition between the different functions required for the various steps of the viral replication cycle. For example, there is evidence that the 5' untranslated region of the HIV-1 genome can form two mutually exclusive secondary structures: a metastable branched multiple-hairpin conformation, which is involved in dimerization and packaging; and a more energetically favourable long-distance interaction conformation, which is involved in transcription and translation. The transition from the long-distance interaction to the branched multiple hairpin conformation is catalysed by the RNA chaperone nucleocapsid protein⁶⁶.

RNA switches can also couple distinct processes within a given step. For example, an RNA switch is used to couple dimerization and selective encapsidation of two copies of the Moloney murine leukaemia virus RNA genome. Dimerization of two RNA genomes induces a shift in the base-pairing pattern within the ψ -RNA packaging signal, which exposes conserved UCUG elements that bind the nucleocapsid protein with high affinity, thereby promoting genome packaging⁶⁷ (Fig. 3e). These elements are base-paired and bind nucleocapsid protein weakly in the monomeric RNA (Fig. 3e).

Functions of tertiary conformational changes

RNA tertiary conformational changes can range from large global changes in the orientation of helices to more subtle local changes in the structure of motifs that are involved in tertiary interactions. These conformational transitions allow RNA molecules to bind adaptively to a wide range of molecular partners and can help to direct the assembly of RNPs.

Polyvalent binding

Some of the first solved structures of RNA–protein complexes revealed a remarkable ability of RNA to undergo adaptive changes in conformation^{2,25} that had the potential to allow the optimization of intermolecular interactions with disparate targets. In a classic example, these conformational changes allow tRNAs to interact with many diverse partners, including ribonuclease P (RNase P), various nucleotide modifying enzymes, tRNA synthetase, EF-Tu, the ribosome and other RNA elements. High-resolution structures of tRNA, tRNA–protein and

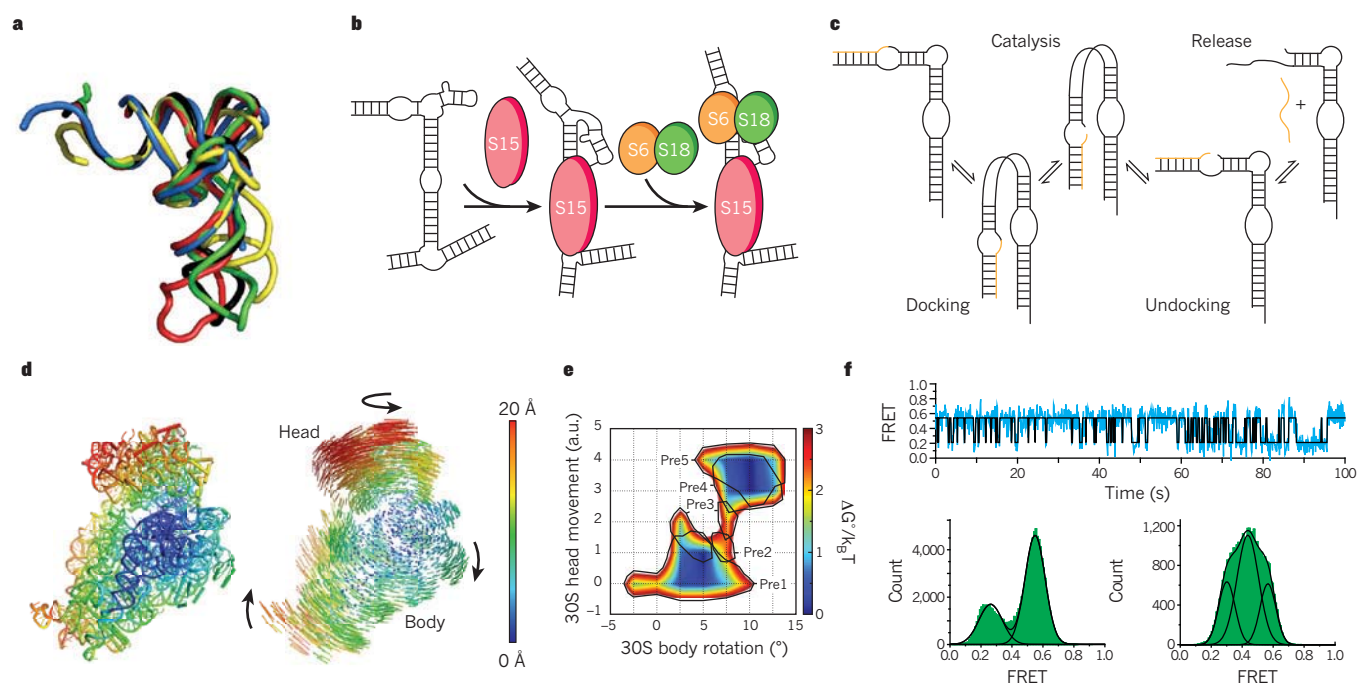


Figure 4 | Functional outputs of tertiary conformational changes. **a**, Different X-ray structures of tRNA^{Phe} in the unbound state (black, PDB ID 1EHZ), in complex with RNaseP (blue, engineered anticodon stem removed, PDB ID 3Q1Q), the ribosome in the P/E state (green, PDB ID 3R8N), isopentenyl-tRNA transferase (red, PDB ID 3FOZ), and phenylalanyl-tRNA synthetase (yellow, PDB ID 1EIY). The structures are superimposed by the acceptor stem. **b**, Hierarchical assembly of the central domain of the 30S ribosomal subunit by successive protein-induced changes in the conformation of 16S rRNA. S15 changes the orientation of the helical domains to favour the binding of S6 and S18. **c**, Enzymatic cycle of the hairpin ribozyme. **d**, Ratcheting motions of the ribosome seen through X-ray crystallography. The degree of 30S subunit atomic displacement between the unratcheted and R₂ ratcheted states with the 50S subunit as a reference (not shown) are colour-coded by Å. Atomic displacement vectors and

arrows (on the right) indicate the direction of the change. Figure reprinted, with permission, from ref. 82. **e**, The free-energy landscape of ribosomal ratcheting, as calculated from subclassification of cryoelectron microscopy particles. Movements of the 30S subunit body and head domains in relation to the 50S subunit are shown in units of degrees and arbitrary units (a.u.), respectively, with corresponding tRNA translocation intermediates (Pre1 and so on) outlined in black. Figure reprinted, with permission, from ref. 84. **f**, Dynamics of the 50S ribosomal L1 stalk monitored by single-molecule fluorescence resonance energy transfer (smFRET). Representative smFRET trace (top) and histogram (bottom left) of the L1 stalk dynamically sampling open and closed conformations in A- and P-site tRNA-bound ribosome complexes. Translocation by EF-G and tRNA occupation of the E- and P-sites causes the L1 stalk conformation to shift dramatically (bottom right). Figure modified, with permission, from ref. 100.

tRNA–RNP complexes show that binding is often accompanied by significant conformational changes, which range from the reorientation of helical domains to finer changes in local structure, all of which optimize intermolecular interactions⁶⁸ (Fig. 4a).

Ordering RNP assembly

RNA tertiary conformational changes that are induced by successive protein-binding events are thought to help direct the order of assembly of complex RNP machines, including the 30S ribosome^{69,70}, the signal recognition particle⁷¹ and telomerase⁷². For example, the binding of ribosomal protein S15 to 16S rRNA initiates the ordered assembly of the central domain in the 30S ribosomal subunit⁷³, and leads to a change in the orientation of helical domains that favours the binding of ribosomal proteins S6 and S18 (ref. 74) (Fig. 4b). Premature binding of S6 and S18 to the unbound 16S rRNA may be disfavoured, in part, because of the entropic penalty that is associated with the partial freezing-out of interhelical motions. Even in the simpler telomerase RNP (consisting of one RNA and two protein components), the binding of the first protein p65 induces a conformational change in the RNA that facilitates the binding of telomerase reverse transcriptase, thereby ordering assembly⁷².

Assembly can also involve coupled protein binding that induces changes in both RNA secondary and tertiary structure. For example, coupled binding of the maturase and Mrs1 protein cofactors to the RNA of the bI3 group-I intron RNP stabilizes both the native tertiary contacts and induces a reorganization of a non-native intermediate secondary structure⁷⁵. Although both Mrs1 dimers and maturase can independently bind and stabilize portions of the bI3 tertiary structure, binding by both proteins is required to induce the secondary structure rearrangement and assembly to the native, active state.

Ribozyme catalysis

Tertiary conformational transitions involving large changes in the orientation of helical arms are often observed in small ribozymes, such as the hairpin and HDV, and are thought to be important for the transition between the different steps of the catalytic cycles. Typically, an undocked (inactive) conformation binds the substrate, promoting the transition into a docked (active) conformation, which is required for catalysis. After catalysis, another undocking transition allows the release of the product (Fig. 4c). The importance of these motions is demonstrated by the fact that the junction motions can accelerate the rate of folding of the active conformation⁷⁶. Similarly, large hinge-like motions of the J2a/b bulge in human telomerase have been proposed to help with dynamic telomere repeat synthesis⁷⁷. A more exceptional example is the *Tetrahymena* group-I ribozyme that has been shown to interconvert between alternative tertiary conformations, which have a range of substrate binding affinities but similar enzymatic activities⁷⁸. The rates of interconversion between these states are slower than the rate of catalysis, implying the existence of multiple native states. Such long-lived heterogeneities have been observed in the tertiary folds of many other RNAs, although some of these may be the result of RNA purification side-products⁷⁹. The atomic level structural differences between these species and the source of the severe heterogeneity are still unknown but they may constitute yet another mechanism used by RNA to define a narrow set of differentiated conformations and this should be an exciting topic for future research.

Protein synthesis

Perhaps the best example of the cell manipulating the intrinsic dynamic landscape of RNA to achieve a desired biological outcome is ribosome catalysis. Large-scale ratcheting motions are required for translation. The small and large subunits reorient with respect to one another through numerous structural intermediates that are driven by changes in the conformation of both the ribosomal RNAs and proteins^{80–84} (Fig. 4d). Data strongly indicate that all of these

intermediates are relatively low-lying energy states that readily interconvert, which has been highlighted by the ability of the ribosome to spontaneously undergo full tRNA retrotranslocation^{84,85} (Fig. 4e). This has led to a ‘Brownian machine’ model of the ribosome, where the ribosome’s functionality is derived in part by its ability to harness thermally driven equilibrium fluctuations and bias them to promote the translation process²² (Fig. 4f).

The cell combines these intrinsic ribosome dynamics with numerous effectors to achieve tight control over the complex transactions that are required by translation. One such transaction is the selection and proofreading of incoming tRNAs that are responsible for the ribosome’s remarkable ability to consistently discriminate between cognate and near- or non-cognate tRNAs, in which small differences between the minihelices of incorrect and correct anticodon–codon pairs will lead to tRNA accommodation or rejection. Here, the formation of a cognate minihelix results in a kinked tRNA structure and triggers a 30S ‘domain closure’ motion^{86–88}. This stabilizes tRNA–ribosome interactions and in turn promotes conformational rearrangements in the EF-Tu protein of the EF-Tu•GTP-tRNA ternary complex that delivered tRNA to the ribosome; this results in EF-Tu•GTP hydrolysis, release of tRNA from EF-Tu and initial tRNA selection^{89,90}. The second proofreading step that follows EF-Tu dissociation is thought to be driven by relaxation of the kinked tRNA. In cognate tRNAs, the strong interactions between the codon and the anticodon cause a bias of tRNA relaxation towards a conformation that is fully accommodated within the A-site. However, for near-cognate tRNAs, which have weak codon–anticodon interactions, the relaxation of the kinked tRNA can occur through other pathways that lead to rejection^{91,92}. Following tRNA accommodation, other factors, including EF-G⁹³, other initiation factors⁹⁴, recycling factors⁹⁵, release factors⁹⁶, and even the identity and acylation state of the tRNA occupying the neighbouring ribosomal P-site⁹⁷, act on the translation process, manipulating the ribosome’s dynamic landscape to drive efficient synthesis of the mRNA-encoded protein.

Owing to the overwhelming complexity of the ribosome, the mechanisms and atomic level details of the many conformational transitions involved in the translation process remain unclear. Among these unresolved questions are how the ribosome’s RNA and protein components cooperate to confer dynamic specificity and robustness on ribosome dynamics⁹⁸. Research into this process is another exciting area of future study, and we can confidently predict that this will be yet another biological system shown to rely heavily on the virtuosity of RNA dynamics.

Outlook

The conventional view that one sequence codes for one structure and one function is being replaced by a dynamic view of RNA as a pre-existing superposition of conformational states that can be resolved into a directed and synchronized motion by dedicated cellular machinery, leading to a broad range of functional outcomes. This makes it all the more important to study RNA dynamics within the complex *in vivo* environment of living cells, an important goal for the future. We also need to increase our basic understanding of RNA dynamic behaviour, even within the simpler *in vitro* environment. It is remarkable that, even for well-studied molecules such as tRNA, there is very little experimental data available regarding the equilibrium fluctuations in tRNA at the atomic level; the same is also true for catalytically important motions in ribozymes. Similarly, little is known about the structure and dynamics of large RNAs, such as eukaryotic mRNAs. This will require the combined development of computational and experimental tools to move towards developing atomic-level movies of RNA in dynamic action within living cells as well as a better predictive understanding of RNA dynamic behaviour. In the meantime, great advances can be made by simply embracing this new dynamic view of RNA and always being on the lookout for another myoglobin. ■

1. Kendrew, J. C. *et al.* A three-dimensional model of the myoglobin molecule obtained by X-ray analysis. *Nature* **181**, 662–666 (1958).
2. Rould, M. A., Perona, J. J., Söll, D. & Steitz, T. A. Structure of *E. coli* glutamyl-tRNA synthetase complexed with tRNA(Gln) and ATP at 2.8 Å resolution. *Science* **246**, 1135–1142 (1989).
3. Pley, H. W., Flaherty, K. M. & McKay, D. B. Three-dimensional structure of a hammerhead ribozyme. *Nature* **372**, 68–74 (1994).
4. Scott, W. G., Finch, J. T. & Klug, A. The crystal structure of an all-RNA hammerhead ribozyme: a proposed mechanism for RNA catalytic cleavage. *Cell* **81**, 991–1002 (1995).
5. Wang, S., Karbstein, K., Peracchi, A., Beigelman, L. & Herschlag, D. Identification of the hammerhead ribozyme metal ion binding site responsible for rescue of the deleterious effect of a cleavage site phosphorothioate. *Biochemistry* **38**, 14363–14378 (1999).
6. Boehr, D. D., Nussinov, R. & Wright, P. E. The role of dynamic conformational ensembles in biomolecular recognition. *Nature Chem. Biol.* **5**, 789–796 (2009).
7. Al-Hashimi, H. M. & Walter, N. G. RNA dynamics: it is about time. *Curr. Opin. Struct. Biol.* **18**, 321–329 (2008).
8. Frauenfelder, H., Sligar, S. G. & Wolynes, P. G. The energy landscapes and motions of proteins. *Science* **254**, 1598–1603 (1991).
9. Cruz, J. A. & Westhof, E. The dynamic landscapes of RNA architecture. *Cell* **136**, 604–609 (2009).
10. Bailor, M. H., Mustoe, A. M., Brooks, C. L. 3rd & Al-Hashimi, H. M. Topological constraints: using RNA secondary structure to model 3D conformation, folding pathways, and dynamic adaptation. *Curr. Opin. Struct. Biol.* **21**, 296–305 (2011).
11. Schultes, E. A., Spasic, A., Mohanty, U. & Bartel, D. P. Compact and ordered collapse of randomly generated RNA sequences. *Nature Struct. Mol. Biol.* **12**, 1130–1136 (2005).
12. Schultes, E. A., Hraber, P. T. & LaBean, T. H. Estimating the contributions of selection and self-organization in RNA secondary structure. *J. Mol. Evol.* **49**, 76–83 (1999).
13. Fürtig, B., Wenter, P., Pitsch, S. & Schwalbe, H. Probing mechanism and transition state of RNA refolding. *ACS Chem. Biol.* **5**, 753–765 (2010).
14. Bailor, M. H., Sun, X. & Al-Hashimi, H. M. Topology links RNA secondary structure with global conformation, dynamics, and adaptation. *Science* **327**, 202–206 (2010).
- This article reports the simple topological constraints that are governed by steric and stereochemical forces severely restrict the allowed orientation of helices across two-way junctions.**
15. Mustoe, A. M., Bailor, M. H., Teixeira, R. M., Brooks, C. L. 3rd & Al-Hashimi, H. M. New insights into the fundamental role of topological constraints as a determinant of two-way junction conformation. *Nucleic Acids Res.* **40**, 892–904 (2012).
16. Chu, V. B. *et al.* Do conformational biases of simple helical junctions influence RNA folding stability and specificity? *RNA* **15**, 2195–2205 (2009).
17. Venditti, V., Clos, L. 2nd, Niccolai, N. & Butcher, S. E. Minimum-energy path for a U6 RNA conformational change involving protonation, base-pair rearrangement and base flipping. *J. Mol. Biol.* **391**, 894–905 (2009).
18. Fourmy, D., Yoshizawa, S. & Puglisi, J. D. Paromomycin binding induces a local conformational change in the A-site of 16S rRNA. *J. Mol. Biol.* **277**, 333–345 (1998).
19. Le, S. Y., Zhang, K. & Maizel, J. V. Jr. RNA molecules with structure dependent functions are uniquely folded. *Nucleic Acids Res.* **30**, 3574–3582 (2002).
20. Stelzer, A. C., Kratz, J. D., Zhang, Q. & Al-Hashimi, H. M. RNA dynamics by design: biasing ensembles towards the ligand-bound state. *Angew. Chem. Int. Ed. Engl.* **49**, 5731–5733 (2010).
21. Shankar, N. *et al.* NMR reveals the absence of hydrogen bonding in adjacent UU and AG mismatches in an isolated internal loop from ribosomal RNA. *Biochemistry* **46**, 12665–12678 (2007).
22. Frank, J. & Gonzalez, R. L., Jr. Structure and dynamics of a processive Brownian motor: the translating ribosome. *Annu. Rev. Biochem.* **79**, 381–412 (2010).
23. Haller, A., Souliere, M. F. & Micura, R. The dynamic nature of RNA as key to understanding riboswitch mechanisms. *Acc. Chem. Res.* **44**, 1339–1348 (2011).
24. Paukstis, P. J., Chen, J. H., Chase, E., Lambowitz, A. M. & Golden, B. L. Structure of a tyrosyl-tRNA synthetase splicing factor bound to a group I intron RNA. *Nature* **451**, 94–97 (2008).
25. Puglisi, J. D., Tan, R., Calnan, B. J., Frankel, A. D. & Williamson, J. R. Conformation of the TAR RNA-arginine complex by NMR spectroscopy. *Science* **257**, 76–80 (1992).
26. Orr, J. W., Hagerman, P. J. & Williamson, J. R. Protein and Mg²⁺-induced conformational changes in the S15 binding site of 16S ribosomal RNA. *J. Mol. Biol.* **275**, 453–464 (1998).
27. Turner, B., Melcher, S. E., Wilson, T. J., Norman, D. G. & Lilley, D. M. Induced fit of RNA on binding the L7Ae protein to the kink-turn motif. *RNA* **11**, 1192–1200 (2005).
28. Falb, M., Amata, I., Gabel, F., Simon, B. & Carlomagno, T. Structure of the K-turn U4 RNA: a combined NMR and SANS study. *Nucleic Acids Res.* **38**, 6274–6285 (2010).
29. Kim, H. D. *et al.* Mg²⁺-dependent conformational change of RNA studied by fluorescence correlation and FRET on immobilized single molecules. *Proc. Natl Acad. Sci. USA* **99**, 4284–4289 (2002).
30. Zacharias, M. & Hagerman, P. J. The influence of symmetric internal loops on the flexibility of RNA. *J. Mol. Biol.* **257**, 276–289 (1996).
31. Zhang, Q., Stelzer, A. C., Fisher, C. K. & Al-Hashimi, H. M. Visualizing spatially correlated dynamics that directs RNA conformational transitions. *Nature* **450**, 1263–1267 (2007).
32. Shajani, Z., Drobny, G. & Varani, G. Binding of U1A protein changes RNA dynamics as observed by ¹³C NMR relaxation studies. *Biochemistry* **46**, 5875–5883 (2007).
33. Bokinsky, G. *et al.* Two distinct binding modes of a protein cofactor with its target RNA. *J. Mol. Biol.* **361**, 771–784 (2006).
34. Bardaro, M. F. Jr., Shajani, Z., Patora-Komisarska, K., Robinson, J. A. & Varani, G. How binding of small molecule and peptide ligands to HIV-1 TAR alters the RNA motional landscape. *Nucleic Acids Res.* **37**, 1529–1540 (2009).
35. Herschlag, D., Khosla, M., Tsuchihashi, Z. & Karpel, R. L. An RNA chaperone activity of non-specific RNA binding proteins in hammerhead ribozyme catalysis. *EMBO J.* **13**, 2913–2924 (1994).
36. Pyle, A. M. & Green, J. B. RNA folding. *Curr. Opin. Struct. Biol.* **5**, 303–310 (1995).
37. Treiber, D. K. & Williamson, J. R. Beyond kinetic traps in RNA folding. *Curr. Opin. Struct. Biol.* **11**, 309–314 (2001).
38. Hirling, H., Scheffner, M., Restle, T. & Stahl, H. RNA helicase activity associated with the human p68 protein. *Nature* **339**, 562–564 (1989).
39. Yang, Q. & Jankowsky, E. ATP- and ADP-dependent modulation of RNA unwinding and strand annealing activities by the DEAD-box protein DED1. *Biochemistry* **44**, 13591–13601 (2005).
40. Will, C. L. & Lührmann, R. Spliceosome structure and function. *Cold Spring Harb. Perspect. Biol.* **3**, a003707 (2011).
41. Kosowski, T. R., Keys, H. R., Quan, T. K. & Ruby, S. W. DExD/H-box Prp5 protein is in the spliceosome during most of the splicing cycle. *RNA* **15**, 1345–1362 (2009).
42. Maeder, C., Kutach, A. K. & Guthrie, C. ATP-dependent unwinding of U4/U6 snRNAs by the Brr2 helicase requires the C terminus of Prp8. *Nature Struct. Mol. Biol.* **16**, 42–48 (2009).
43. Schwer, B. A conformational rearrangement in the spliceosome sets the stage for Prp22-dependent mRNA release. *Mol. Cell* **30**, 743–754 (2008).
44. Bhaskaran, H. & Russell, R. Kinetic redistribution of native and misfolded RNAs by a DEAD-box chaperone. *Nature* **449**, 1014–1018 (2007).
45. Winkler, W., Nahvi, A. & Breaker, R. R. Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature* **419**, 952–956 (2002).
- This article reports the discovery of an RNA switch in the 5' untranslated region of bacterial mRNA that regulates gene expression in response to ligands without assistance from proteins.**
46. Cromie, M. J., Shi, Y., Latifi, T. & Groisman, E. A. An RNA sensor for intracellular Mg²⁺. *Cell* **125**, 71–84 (2006).
47. Nechooshtan, G., Elgrably-Weiss, M., Sheaffer, A., Westhof, E. & Altuvia, S. A pH-responsive riboregulator. *Genes Dev.* **23**, 2650–2662 (2009).
48. Greenleaf, W. J., Frieda, K. L., Foster, D. A., Woodside, M. T. & Block, S. M. Direct observation of hierarchical folding in single riboswitch aptamers. *Science* **319**, 630–633 (2008).
49. Mandal, M. *et al.* A glycine-dependent riboswitch that uses cooperative binding to control gene expression. *Science* **306**, 275–279 (2004).
50. Sudarsan, N. *et al.* Tandem riboswitch architectures exhibit complex gene control functions. *Science* **314**, 300–304 (2006).
51. Lee, E. R., Baker, J. L., Weinberg, Z., Sudarsan, N. & Breaker, R. R. An allosteric self-splicing ribozyme triggered by a bacterial second messenger. *Science* **329**, 845–848 (2010).
52. Ferre-D'Amare, A. R., Zhou, K. & Doudna, J. A. Crystal structure of a hepatitis delta virus ribozyme. *Nature* **395**, 567–574 (1998).
53. Ke, A., Zhou, K., Ding, F., Cate, J. H. D. & Doudna, J. A. A conformational switch controls hepatitis delta virus ribozyme catalysis. *Nature* **429**, 201–205 (2004).
- This article reports a significant local conformational change in the active site of the HDV ribozyme is observed post-cleavage and is associated with ejection of the substrate and a catalytically critical divalent metal ion.**
54. Harris, D. A., Rueda, D. & Walter, N. G. Local conformational changes in the catalytic core of the *trans*-acting hepatitis delta virus ribozyme accompany catalysis. *Biochemistry* **41**, 12051–12061 (2002).
55. Lamanna, A. C. & Karbstein, K. An RNA conformational switch regulates pre-18S rRNA cleavage. *J. Mol. Biol.* **405**, 3–17 (2011).
56. Nocker, A. *et al.* A mRNA-based thermosensor controls expression of rhizobial heat shock genes. *Nucleic Acids Res.* **29**, 4800–4807 (2001).
57. Johansson, J. *et al.* An RNA thermosensor controls expression of virulence genes in *Listeria monocytogenes*. *Cell* **110**, 551–561 (2002).
58. Watts, J. M. *et al.* Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature* **460**, 711–716 (2009).
59. Grundy, F. J., Winkler, W. C. & Henkin, T. M. tRNA-mediated transcription antitermination *in vitro*: codon-anticodon pairing independent of the ribosome. *Proc. Natl Acad. Sci. USA* **99**, 11121–11126 (2002).
60. Babitzke, P. & Yanofsky, C. Reconstitution of *Bacillus subtilis* *trp* attenuation *in vitro* with TRAP, the *trp* RNA-binding attenuation protein. *Proc. Natl Acad. Sci. USA* **90**, 133–137 (1993).
61. Diaz-Toledano, R., Ariza-Mateos, A., Birk, A., Martinez-Garcia, B. & Gomez, J. *In vitro* characterization of a miR-122-sensitive double-helical switch element in the 5' region of hepatitis C virus RNA. *Nucleic Acids Res.* **37**, 5498–5510 (2009).
62. Ray, P. S. *et al.* A stress-responsive RNA switch regulates VEGFA expression. *Nature* **457**, 915–919 (2009).
- This article reports that the 3' untranslated region of human VEGFA mRNA undergoes a binary conformational switch in response to inflammatory and hypoxic protein stress signals to regulate VEGFA expression.**
63. Cheah, M. T., Wachter, A., Sudarsan, N. & Breaker, R. R. Control of alternative RNA splicing and gene expression by eukaryotic riboswitches. *Nature* **447**, 497–500 (2007).

This article reports a secondary structural change in a eukaryotic thiamine pyrophosphate riboswitch regulates gene expression through the control of alternative splicing.

64. Kedde, M. *et al.* A Pumilio-induced RNA structure switch in p27-3' untranslated region controls miR-221 and miR-222 accessibility. *Nature Cell Biol.* **12**, 1014–1020 (2010).
65. Casey, J. L. Control of ADAR1 editing of hepatitis delta virus RNAs. *Curr. Top. Microbiol. Immunol.* **353**, 123–143 (2012).
66. Abbink, T. E., Ooms, M., Haasnoot, P. C. & Berkhout, B. The HIV-1 leader RNA conformational switch regulates RNA dimerization but does not regulate mRNA translation. *Biochemistry* **44**, 9058–9066 (2005).
67. Miyazaki, Y. *et al.* An RNA structural switch regulates diploid genome packaging by Moloney murine leukemia virus. *J. Mol. Biol.* **396**, 141–152 (2010).
- This article reports that dimerization of the 5' untranslated region of the Moloney murine leukaemia virus results in a secondary structural change that promotes genome packaging.**
68. Gieger, R. Toward a more complete view of tRNA biology. *Nature Struct. Mol. Biol.* **15**, 1007–1014 (2008).
69. Mulder, A. M. *et al.* Visualizing ribosome biogenesis: parallel assembly pathways for the 30S subunit. *Science* **330**, 673–677 (2010).
70. Adilakshmi, T., Bellur, D. L. & Woodson, S. A. Concurrent nucleation of 16S folding and induced fit in 30S ribosome assembly. *Nature* **455**, 1268–1272 (2008).
71. Menichelli, E., Isel, C., Oubridge, C. & Nagai, K. Protein-induced conformational changes of RNA during the assembly of human signal recognition particle. *J. Mol. Biol.* **367**, 187–203 (2007).
72. Stone, M. D. *et al.* Stepwise protein-mediated RNA folding directs assembly of telomerase ribonucleoprotein. *Nature* **446**, 458–461 (2007).
73. Held, W. A., Ballou, B., Mizushima, S. & Nomura, M. Assembly mapping of 30S ribosomal proteins from *Escherichia coli*. Further studies. *J. Biol. Chem.* **249**, 3103–3111 (1974).
74. Agalarov, S. C., Prasad, G. S., Funke, P. M., Stout, C. D. & Williamson, J. R. Structure of the S15, S6, S18-rRNA complex: assembly of the 30S ribosome central domain. *Science* **288**, 107–112 (2000).
75. Duncan, C. D. & Weeks, K. M. Nonhierarchical ribonucleoprotein assembly suggests a strain-propagation model for protein-facilitated RNA folding. *Biochemistry* **49**, 5418–5425 (2010).
76. Wilson, T. J., Nahas, M., Ha, T. & Lilley, D. M. Folding and catalysis of the hairpin ribozyme. *Biochem. Soc. Trans.* **33**, 461–465 (2005).
77. Zhang, Q., Kim, N. K., Peterson, R. D., Wang, Z. & Feigon, J. Structurally conserved five nucleotide bulge determines the overall topology of the core domain of human telomerase RNA. *Proc. Natl Acad. Sci. USA* **107**, 18761–18768 (2010).
78. Solomatina, S. V., Greenfield, M., Chu, S. & Herschlag, D. Multiple native states reveal persistent ruggedness of an RNA folding landscape. *Nature* **463**, 681–684 (2010).
- This article reports the observation of slowly interconverting catalytically active states in a ribozyme, thereby establishing the coexistence of multiple native states.**
79. Greenfield, M., Solomatina, S. V. & Herschlag, D. Removal of covalent heterogeneity reveals simple folding behavior for P4–P6 RNA. *J. Biol. Chem.* **286**, 19872–19879 (2011).
80. Frank, J. & Agrawal, R. K. A ratchet-like inter-subunit reorganization of the ribosome during translocation. *Nature* **406**, 318–322 (2000).
81. Valle, M. *et al.* Locking and unlocking of ribosomal motions. *Cell* **114**, 123–134 (2003).
82. Zhang, W., Dunkle, J. A. & Cate, J. H. Structures of the ribosome in intermediate states of ratcheting. *Science* **325**, 1014–1017 (2009).
83. Ratje, A. H. *et al.* Head swivel on the ribosome facilitates translocation by means of intra-subunit tRNA hybrid sites. *Nature* **468**, 713–716 (2010).
84. Fischer, N., Konevega, A. L., Wintermeyer, W., Rodnina, M. V. & Stark, H. Ribosome dynamics and tRNA movement by time-resolved electron cryomicroscopy. *Nature* **466**, 329–333 (2010).

This article demonstrates the cryo-electron microscopy observation of thermally driven tRNA retrotranslocation on the ribosome.

85. Shoji, S., Walker, S. E. & Fredrick, K. Reverse translocation of tRNA in the ribosome. *Mol. Cell* **24**, 931–942 (2006).
86. Ogle, J. M., Murphy, F. V., Tarry, M. J. & Ramakrishnan, V. Selection of tRNA by the ribosome requires a transition from an open to a closed form. *Cell* **111**, 721–732 (2002).
87. Valle, M. *et al.* Incorporation of aminoacyl-tRNA into the ribosome as seen by cryo-electron microscopy. *Nature Struct. Mol. Biol.* **10**, 899–906 (2003).
88. Lee, T. H., Blanchard, S. C., Kim, H. D., Puglisi, J. D. & Chu, S. The role of fluctuations in tRNA selection by the ribosome. *Proc. Natl Acad. Sci. USA* **104**, 13661–13665 (2007).
89. Schmeing, T. M. *et al.* The crystal structure of the ribosome bound to EF-Tu and aminoacyl-tRNA. *Science* **326**, 688–694 (2009).
90. Voorhees, R. M., Schmeing, T. M., Kelley, A. C. & Ramakrishnan, V. The mechanism for activation of GTP hydrolysis on the ribosome. *Science* **330**, 835–838 (2010).
91. Pape, T., Wintermeyer, W. & Rodnina, M. V. Conformational switch in the decoding region of 16S rRNA during aminoacyl-tRNA selection on the ribosome. *Nature Struct. Mol. Biol.* **7**, 104–107 (2000).
92. Blanchard, S. C., Gonzalez, R. L., Kim, H. D., Chu, S. & Puglisi, J. D. tRNA selection and kinetic proofreading in translation. *Nature Struct. Mol. Biol.* **11**, 1008–1014 (2004).
- This important single-molecule FRET study directly observes the dynamics of tRNA initial selection and proofreading by the ribosome.**
93. Fei, J. *et al.* Allosteric collaboration between elongation factor G and the ribosomal L1 stalk directs tRNA movements during translation. *Proc. Natl Acad. Sci. USA* **106**, 15702–15707 (2009).
94. Blaha, G., Stanley, R. E. & Steitz, T. A. Formation of the first peptide bond: the structure of EF-P bound to the 70S ribosome. *Science* **325**, 966–970 (2009).
95. Dunkle, J. A. *et al.* Structures of the bacterial ribosome in classical and hybrid states of tRNA binding. *Science* **332**, 981–984 (2011).
96. Laurberg, M. *et al.* Structural basis for translation termination on the 70S ribosome. *Nature* **454**, 852–857 (2008).
97. Cornish, P. V., Ermolenko, D. N., Noller, H. F. & Ha, T. Spontaneous intersubunit rotation in single ribosomes. *Mol. Cell* **30**, 578–588 (2008).
98. Tama, F., Valle, M., Frank, J. & Brooks, C. L. 3rd. Dynamic reorganization of the functionally active ribosome explored by normal mode analysis and cryo-electron microscopy. *Proc. Natl Acad. Sci. USA* **100**, 9319–9323 (2003).
99. Green, N. J., Grundy, F. J. & Henkin, T. M. The T box mechanism: tRNA as a regulatory molecule. *FEBS Lett.* **584**, 318–324 (2010).
100. Cornish, P. V. *et al.* Following movement of the L1 stalk between three functional states in single ribosomes. *Proc. Natl Acad. Sci. USA* **106**, 2571–2576 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements E.A.D. and J.C. contributed equally to this Review. We thank C. Eichhorn and Q. Zhang for their input and assistance in the preparation of figures, and S. Butcher and S. Serganov for their comments on this Review. A.M.M. is supported by an NSF graduate research fellowship. The authors gratefully acknowledge the Michigan Economic Development Cooperation and the Michigan Technology Tri-Corridor for their support in the purchase of a 600 MHz spectrometer. This work was supported by the US National Institutes of Health (R01 AI066975 and R01 GM089846) and the US National Science Foundation (NSF Career Award CHE-0918817).

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcomed to comment on the online version of this article at www.nature.com/nature. Correspondence should be addressed to author H.M.A.H. (hashimi@umich.edu).

RNA-guided genetic silencing systems in bacteria and archaea

Blake Wiedenheft^{1,2†}, Samuel H. Sternberg³ & Jennifer A. Doudna¹⁻⁴

Clustered regularly interspaced short palindromic repeat (CRISPR) are essential components of nucleic-acid-based adaptive immune systems that are widespread in bacteria and archaea. Similar to RNA interference (RNAi) pathways in eukaryotes, CRISPR-mediated immune systems rely on small RNAs for sequence-specific detection and silencing of foreign nucleic acids, including viruses and plasmids. However, the mechanism of RNA-based bacterial immunity is distinct from RNAi. Understanding how small RNAs are used to find and destroy foreign nucleic acids will provide new insights into the diverse mechanisms of RNA-controlled genetic silencing systems.

Bacteria and archaea are the most diverse and abundant organisms on the planet, thriving in habitats that range from hot springs to humans. However, viruses outnumber their microbial hosts in every ecological setting, and the selective pressures imposed by these rapidly evolving parasites has driven the diversification of microbial defence systems¹⁻³. Historically, our understanding of antiviral immunity in bacteria has focused on restriction-modification systems, abortive-phage phenotypes, toxin-antitoxins and other innate defence systems^{4,5}. More recently, bioinformatic, genetic and biochemical studies have revealed that many prokaryotes use an RNA-based adaptive immune system to target and destroy genetic parasites (reviewed in refs 6-12). Such adaptive immunity, previously thought to occur only in eukaryotes, provides an example of RNA-guided destruction of foreign genetic material by a process that is distinct from RNA interference (RNAi) (Fig. 1).

In response to viral and plasmid challenges, bacteria and archaea integrate short fragments of foreign nucleic acid into the host chromosome at one end of a repetitive element known as CRISPR (clustered regularly interspaced short palindromic repeat)¹³⁻¹⁵. These repetitive loci serve as molecular 'vaccination cards' by maintaining a genetic record of prior encounters with foreign transgressors. CRISPR loci are transcribed, and the long primary transcript is processed into a library of short CRISPR-derived RNAs (crRNAs)¹⁶⁻²¹ that each contain a sequence complementary to a previously encountered invading nucleic acid. Each crRNA is packaged into a large surveillance complex that patrols the intracellular environment and mediates the detection and destruction of foreign nucleic acid targets^{15,22-27}.

CRISPRs were originally identified in the *Escherichia coli* genome in 1987, when they were described as an unusual sequence element consisting of a series of 29-nucleotide repeats separated by unique 32-nucleotide 'spacer' sequences²⁸. Repetitive sequences with a similar repeat-spacer-repeat pattern were later identified in phylogenetically diverse bacterial and archaeal genomes, but the function of these repeats remained obscure until many spacer sequences were recognized as being identical to viral and plasmid sequences²⁹⁻³¹. This observation led to the hypothesis that CRISPRs provide a genetic memory of infection²⁹, and the detection of short CRISPR-derived RNA transcripts suggested that there may be functional similarities between CRISPR-based immunity and RNAi^{30,32}. In this Insight, we review three stages of CRISPR-based adaptive immunity and compare mechanistic aspects of these immune systems to other RNA-guided genetic silencing pathways.

Architecture and composition of CRISPR loci

The defining feature of CRISPR loci is a series of direct repeats (approximately 20-50 base pairs) separated by unique spacer sequences of a similar length^{11,33,34} (Fig. 2). The repeat sequences within a CRISPR locus are conserved, but repeats in different CRISPR loci can vary in both sequence and length. In addition, the number of repeat-spacer units in a CRISPR locus varies widely within and among organisms³⁵.

The sequence diversity of these repetitive loci initially limited their detection and obscured their relationship, but computational methods have been developed for detecting repeat patterns rather than related sequences^{33,34,36-38}. One of the first-generation pattern-recognition algorithms identified the repeat-spacer-repeat architecture in phylogenetically diverse bacterial and archaeal genomes, but related structures were not identified in eukaryotic chromosomes³⁹. Comparative analyses of the sequences adjacent to the CRISPR loci have revealed an (A+T)-rich 'leader' sequence that has been shown to serve as a promoter element for CRISPR transcription³⁹⁻⁴². In addition to the leader sequence, Jansen *et al.*³⁹ identified a set of four CRISPR-associated (*cas*) genes known as *cas1-4* that are found exclusively in genomes containing CRISPRs. Based on sequence similarity to proteins of known function, Cas3 was predicted to be a helicase and Cas4 a RecB-like exonuclease³⁹.

Subsequent bioinformatic analyses have shown that CRISPR loci are flanked by a large number of extremely diverse *cas* genes^{32,43}. The *cas1* gene is a common component of all CRISPR systems, and phylogenetic analyses of Cas1 sequences indicate there are several versions of the CRISPR system. Providing additional evidence for the classification of distinct CRISPR types, neighbourhood analysis has identified conserved arrangements of between four and ten *cas* genes that are found in association with CRISPR loci harbouring specific repeat sequences³⁵.

These distinct immune systems have been divided into three major CRISPR types on the basis of gene conservation and locus organization¹⁰. More than one CRISPR type is often found in a single organism, indicating that these systems are probably mutually compatible and could share functional components¹⁰. Despite the variation in number and diversity of *cas* genes, the distinguishing feature of all type I systems is that they encode a *cas3* gene. The Cas3 protein contains an N-terminal HD phosphohydrolase domain and a C-terminal helicase domain^{32,39,43,44}. In some type I systems, the Cas3 nuclease and helicase domains are encoded by separate genes (*cas3'* and *cas3*, respectively), but in each case they are thought to participate in degrading foreign nucleic acids^{22,44-46} (Fig. 2).

¹Howard Hughes Medical Institute, 4000 Jones Bridge Road, Chevy Chase, Maryland 20815-6789, USA. ²Department of Molecular and Cell Biology, University of California, Berkeley, California 94720, USA. ³Department of Chemistry, University of California, Berkeley, California 94720, USA. ⁴Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA.

[†]Present address: Department of Immunology and Infectious Diseases, Montana State University, Bozeman, Montana, USA

Type II CRISPR systems consist of just four *cas* genes, one of which is always *cas9* (formerly referred to as *csn1*). *Cas9* is a large protein that includes both a RuvC-like nuclease domain and an HNH nuclease domain. Studies in *Streptococcus pyogenes* and *Streptococcus thermophilus* have indicated that *Cas9* may participate in both CRISPR RNA processing and target destruction^{14,15,17}. Two variations of the type III system have been identified (known as III-A and III-B). This division is supported by the functional differences reported in *Staphylococcus epidermidis* and *Pyrococcus furiosus*^{47,48}. The immune system in *S. epidermidis* (type III-A) targets plasmid DNA *in vivo*, whereas the purified components of the type III-B system in *P. furiosus* have been found to cleave only single-stranded RNA substrates *in vitro*. The functional distinction between these two closely related systems suggests there could be other mechanistic differences between the distinct CRISPR subtypes.

Integration of new information into CRISPR loci

Acquisition of foreign DNA is the first step of CRISPR-mediated immunity (Fig. 2 and 3). During this stage, a short segment of DNA from an invading virus or plasmid (known as the protospacer) is integrated preferentially at the leader end of the CRISPR locus^{14,15}. Although metagenomic studies performed on environmental samples indicate that CRISPRs evolve rapidly in dynamic equilibrium with resident phage populations^{13,49,50}, the type II system in *S. thermophilus* is currently the only CRISPR system that has been shown to robustly acquire new phage or plasmid sequences in a pure culture. Phage-challenge experiments in *S. thermophilus* have indicated that a small proportion of the cells in a population will typically incorporate a single virus-derived sequence at the leader end of a CRISPR locus^{14,15,51,52}. The CRISPR-repeat sequence is duplicated for each new spacer sequenced added, thus maintaining the repeat-spacer-repeat architecture. Although the mechanism of spacer integration and replication of the repeat sequence is still unknown, studies in *S. thermophilus* and *E. coli* have indicated that several *Cas* proteins are involved in the process^{14,15,22,53}. Mutational analysis of the

cas genes in *S. thermophilus* demonstrated that *csn2* (previously known as *cas7*) is required for new spacer sequence acquisition¹⁴. This gene is not conserved in other CRISPR types, which suggests that either the mechanism of adaptation in *S. thermophilus* is distinct from the other types or that there are functional orthologues of *Csn2* in other systems. Furthermore, gene deletion experiments in both *S. thermophilus* and *E. coli* have shown that neither *cas1* nor *cas2* genes are required for CRISPR RNA processing or targeted interference^{22,53,54}. These genetic studies suggest a role for *Cas1* and *Cas2* in the integration of foreign DNA into the CRISPR.

The role of *Cas1* in CRISPR-mediated immunity is still uncertain; however, biochemical and structural data indicate a function for *Cas1* in new-spacer-sequence acquisition^{54–56}. *Cas1* proteins from *Pseudomonas aeruginosa*⁵⁶, *E. coli*⁵⁴ and *Sulfolobus solfataricus*⁵⁵ have been purified and studied biochemically. The *Cas1* protein from *S. solfataricus* has been shown to bind nucleic acids with high affinity (K_d ranging from 20 to 50 nM), but without sequence preference⁵⁵. The *Cas1* protein from *E. coli* also binds to DNA with a preference for mismatched or abasic substrates⁵⁷. This observation is consistent with a recent study showing a physical and genetic interaction between *E. coli* *Cas1* and several proteins associated with DNA replication and repair⁵⁴.

Activity assays with *Cas1* from *P. aeruginosa* and *E. coli* indicate that *Cas1* is a metal-dependent nuclease. The *Cas1* protein from *P. aeruginosa* is a DNA-specific nuclease, whereas the *Cas1* protein from *E. coli* had a nuclease activity on a wider range of nucleic acid substrates^{54,56}. These *in vitro* assays suggest that *Cas1* proteins interact with nucleic acids in a non-sequence-specific manner.

Crystal structures for five different *Cas1* proteins are currently available (Protein Data Bank (PDB) identifiers: 3GOD, 3NKD, 3LFX, 3PV9 and 2YZS)^{54,56}. Although the amino acid sequences for these proteins are extremely diverse (less than 15% sequence identity), their tertiary and quaternary structures are similar. All *Cas1* proteins seem to share a two-domain architecture consisting of an N-terminal β -strand domain and a

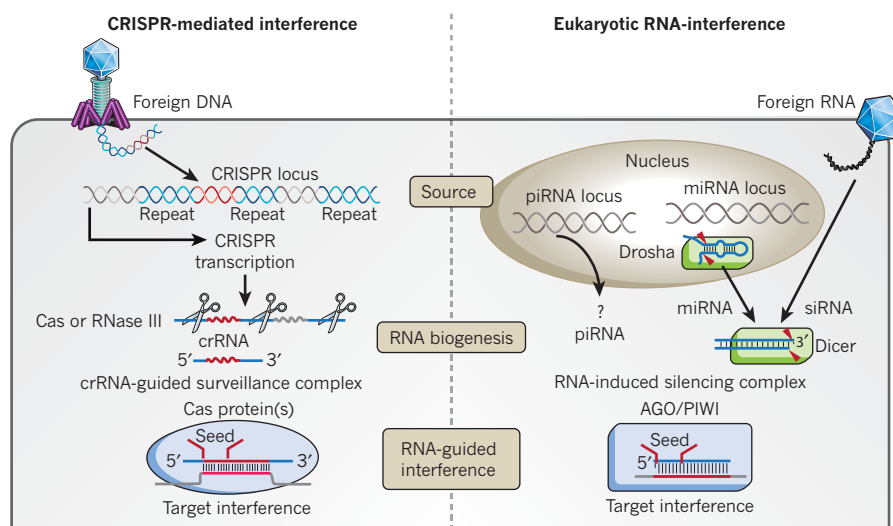


Figure 1 | Parallels and distinctions between CRISPR RNA-guided silencing systems and RNAi. CRISPR systems and RNAi recognize long RNA precursors that are processed into small RNAs, which act as sequence-specific guides for targeting complementary nucleic acids. In CRISPR systems, foreign DNA is integrated into the CRISPR locus, and long transcripts from these loci are processed by a CRISPR-associated (*Cas*) or RNase III family nuclease^{16–21,64}. The short CRISPR-derived RNAs (crRNAs) assemble with *Cas* proteins into large surveillance complexes that target destruction of invading genetic material^{15,22,24–27,48}. In some eukaryotes, long double-stranded RNAs are recognized as foreign, and a specialized RNase III family endoribonuclease (*Dicer*) cleaves these RNAs into short-interfering RNAs (siRNAs) that guide the immune system to invading RNA viruses⁷⁶. PIWI-interacting RNAs (piRNAs) are transcribed from repetitive clusters in the genome that often contain many copies of retrotransposons and primarily

act by restricting transposon mobility^{76–78}. The biogenesis of piRNAs is not yet fully understood. MicroRNAs (miRNAs) are also encoded on the chromosome, and primary miRNA transcripts form stable hairpin structures that are sequentially processed (shown by red triangles) by two RNase III family endoribonucleases (*Drosha* and *Dicer*)⁷⁹. miRNAs do not participate in genome defence but are major regulators of endogenous gene expression⁸⁰. Like crRNAs, eukaryotic piRNAs, siRNAs and miRNAs associate with proteins that facilitate complementary interactions with invading nucleic acid targets^{27,60,69,79}. In eukaryotes, the Argonaute proteins pre-order the 5' region of the guide RNA into a helical configuration, reducing the entropy penalty of interactions with target RNAs⁶⁹. This high-affinity binding site, called the 'seed' sequence, is essential for target sequence interactions. Recent studies indicate that the CRISPR system may use a similar seed-binding mechanism for enhancing target sequence interactions^{26,27,53,60}.

C-terminal α -helical domain (Fig. 3). The C-terminal domain contains a conserved divalent metal-ion binding site, and alanine substitutions of the metal-coordinating residues inhibit Cas1-catalysed DNA degradation^{54,56}. The metal ion is surrounded by a cluster of basic residues that form a strip of positive charge across the surface of the C-terminal domain. This positively charged surface may serve as an electrostatic snare to position nucleic-acid substrates near the catalytic metal ions⁵⁶ (Fig. 3). The Cas1 protein forms a stable homodimer that is formed through interactions between the two β -strand domains, which are related by a pseudo-two-fold axis of symmetry^{54,56}. This organization creates a saddle-like structure that can be modelled onto double-stranded DNA without steric clashing. β -hairpins, one from each of the two symmetrically related molecules, hang on opposite faces of the double-stranded DNA (like stirrups on a

saddle). Although this feature of the Cas1 structure did not initially stand out as a potential DNA-binding site, comparative analysis of the available Cas1 structures reveals a conserved set of positively charged residues along each of the β -hairpins that could contact the phosphate backbone. The two β -hairpins, which are symmetrically related, might participate in sequence-specific interactions with the CRISPR repeat, whereas the large positively charged surface on the C-terminal α -helical domain could account for the high-affinity, non-sequence-specific interactions that have been observed *in vitro*.

In spite of these structural studies and biochemical results, it is still only possible to speculate on the role of Cas1 in the integration of new spacer sequences, and many steps associated with the integration process still need to be explained. For example, new spacer sequences are inserted

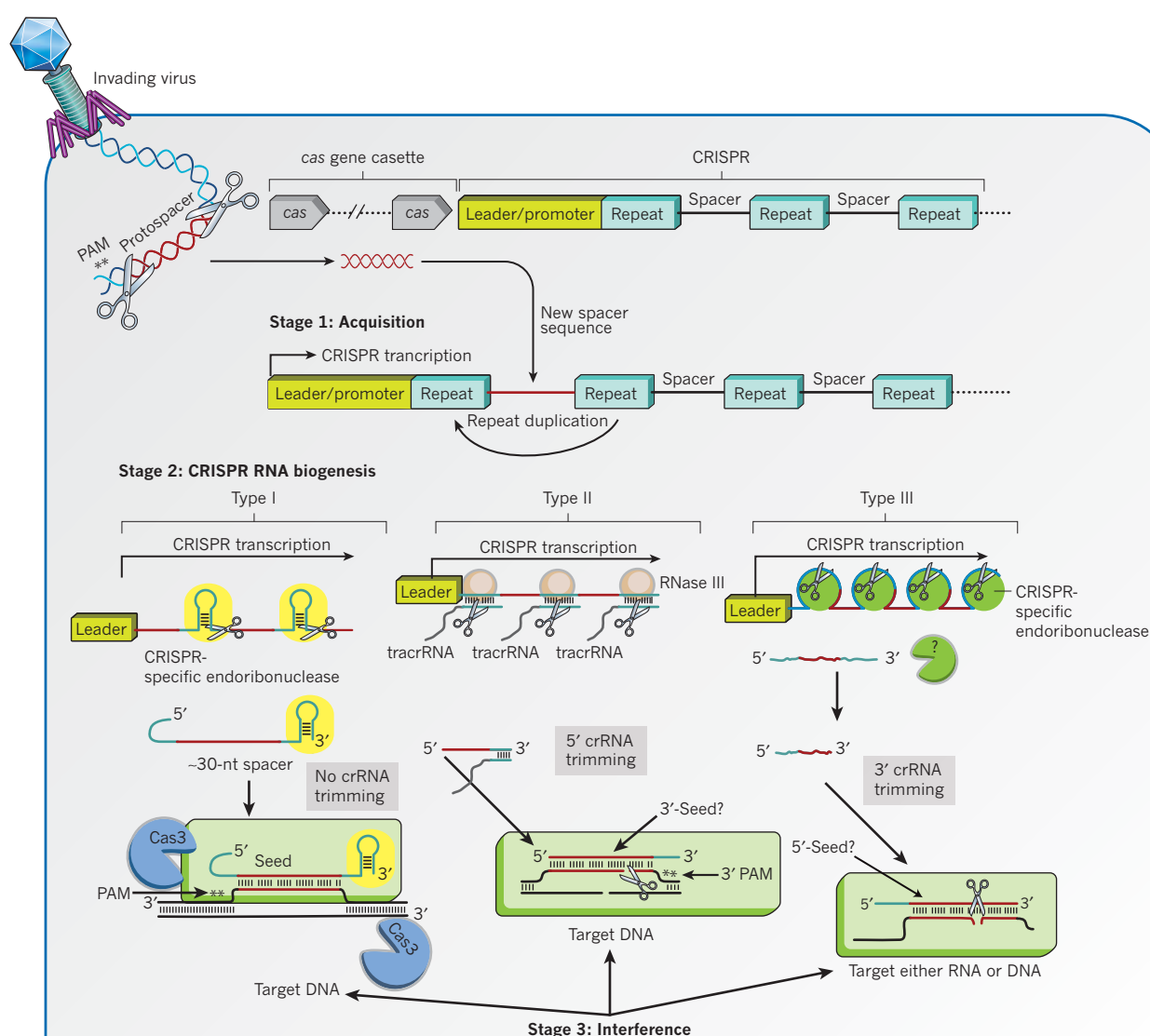


Figure 2 | Diversity of CRISPR-mediated adaptive immune systems in bacteria and archaea. A diverse set of CRISPR-associated (*cas*) genes (grey arrows) encode proteins required for new spacer sequence acquisition (Stage 1), CRISPR RNA biogenesis (Stage 2) and target interference (Stage 3). Each CRISPR locus consists of a series of direct repeats separated by unique spacer sequences acquired from invading genetic elements (protospacers). Protospacers are flanked by a short motif called the protospacer adjacent motif (PAM, **) that is located on the 5' (type I) or 3' (type II) side in foreign DNA^{10,51,52,59,67}. Long CRISPR transcripts are processed into short crRNAs by distinct mechanisms. In type I and III systems, a CRISPR-specific endoribonuclease (yellow ovals and green circles, respectively) cleaves 8 nucleotides upstream of each spacer sequence^{16,18–21,64}. In type II systems, the repeat sequence on the 3'

end of the crRNA is trimmed by an unknown mechanism (green pacman, right). In type II systems, a *trans*-acting antisense RNA (tracrRNA) with complementarity to the CRISPR RNA repeat sequence forms an RNA duplex that is recognized and cleaved by cellular RNase III (brown ovals)¹⁷. This cleavage intermediate is further processed at the 5' end resulting in a mature, approximately 40-nucleotide crRNA with an approximately 20-nucleotide 3'-handle. In each system, the mature crRNA associates with one or more Cas proteins to form a surveillance complex (green rectangles). Type I systems encode a Cas3 nuclease (blue pacman), which may be recruited to the surveillance complex following target binding^{24,27,44}. A short high-affinity binding site called a seed-sequence has been identified in some type I systems^{27,60}, and genetic experiments suggest that type II systems have a seed sequence located at the 3' end of the crRNA spacer sequence⁵³.

preferentially at the leader end of the CRISPR, but the mechanism of leader end recognition is unknown. One simple model suggests that the leader sequence contains a recognition element that recruits the integration machinery. It is equally possible that integration relies on single-stranded regions of the CRISPR DNA that are made available during transcription. Transcription-associated recombination is involved in genome stability⁵⁸, and a mechanism that couples integration together with transcription would link the process of adaptation to CRISPR RNA expression, ensuring that spacers from the most recent virus or plasmid are transcribed first.

The integration machinery must be able to distinguish foreign DNA from that of the host genome. The molecular cues that are involved in the distinction of 'self' from 'non-self' are still unknown, but sequencing of CRISPR loci following phage challenge suggests that spacer sequences are not selected at random^{15,29,51,52,59,60}. Mapping spacer sequences onto viral genomes reveals a short sequence motif proximal to the protospacer, which is referred to as the protospacer adjacent motif (PAM). PAM sequences are only a few nucleotides long, and the precise sequence varies depending on the CRISPR system type⁵⁹. This variation suggests that one or more of the Cas proteins associated with each immune system is involved in PAM recognition, but the mechanism governing this specificity is unknown.

CRISPR RNA biogenesis

Spacer acquisition is the first step of immunization, but successful protection from bacteriophage or plasmid challenge requires the CRISPR to be transcribed and processed into short CRISPR-derived RNAs (crRNAs). crRNAs were first detected by small RNA profiling in *Archaeoglobus fulgidus*⁶¹ and *S. solfataricus*⁶². Northern-blot analysis using probes against the repeat sequence of the CRISPR revealed a 'ladder-like' pattern of RNA consistent with a long precursor CRISPR RNA transcript (pre-crRNA) that was processed at approximately 60-nucleotide intervals. In fact, the 3' ends of cloned crRNAs were mapped to the middle of the CRISPR repeat⁶¹, which suggested that the repeat sequence was recognized and cleaved.

The need for crRNAs in CRISPR-mediated defence was demonstrated initially by investigation of a CRISPR-specific endoribonuclease in *E. coli* called Cas6e (formerly known as Cse3 or CasE)²². Cas6e specifically binds and cleaves within each repeat sequence of the long pre-crRNA, resulting in a library of crRNAs that each contain a unique spacer sequence flanked by fragments of the adjacent repeats. Mutation of a conserved histidine blocks crRNA biogenesis and leaves the cell susceptible to phage infection²².

The Cas6e protein consists of a double ferredoxin-like fold that selectively associates with specific RNA repeats and does not associate with DNA or CRISPR RNAs containing a non-cognate repeat sequence^{18,20,22,63} (Fig. 4). Crystal structures of Cas6e bound to a CRISPR RNA repeat reveal a combination of sequence- and structure-specific interactions that explain the molecular mechanism of substrate recognition^{18,20}. The repeat sequence of the *E. coli* CRISPR is partially palindromic, and the RNA forms a stable (approximately 20-nucleotide) stem loop^{22,35}. A positively charged β -hairpin in Cas6e interacts with the major groove of the RNA duplex, which positions the 3' strand of the crRNA stem along a conserved, positively charged cleft on one face of the protein^{18,20} (Fig. 4). RNA binding induces a conformational change that disrupts the bottom base pair of the stem and positions the scissile phosphate within the enzyme active site for site-specific cleavage²⁰. CRISPR RNA cleavage occurs 8 nucleotides upstream of the spacer sequence, which results in 61-nucleotide mature crRNAs consisting of a 32-nucleotide spacer flanked by 8 nucleotides of the repeat sequence on the 5' end (known as the 5'-handle) and 21 nucleotides of the remaining repeat sequence on the 3' end (Fig. 4). Cas6e remains tightly bound to the 3' stem-loop²⁰ and may serve as a nucleation point for assembly of a large effector complex, Cascade (CRISPR-associated complex for antiviral defence), that is required for phage silencing in the next stage of the immune system^{22,24,26} (discussed later).

Crystal structures of CRISPR-specific endoribonucleases from two other immune systems offer additional insights into the co-evolutionary relationship between these specialized enzymes and their cognate RNAs^{16,19,21} (Fig. 4). In *P. aeruginosa*, Cas6f (previously known as Csy4) specifically binds and cleaves the CRISPR-RNA-repeat 8 nucleotides upstream of the spacer sequence, which leaves a similar 8-nucleotide 5'-handle on mature crRNAs¹⁹. The co-crystal structure of Cas6f bound to its cognate RNA reveals interesting parallels between the method of RNA binding used by Cas6f and Cas6e^{18–20}. Like Cas6e, the *P. aeruginosa* Cas6f protein recognizes the sequence and shape of a stable stem-loop in the crRNA repeat sequence by interacting extensively with the major groove of the double-stranded RNA. However, the structural elements responsible for this interaction are distinct between the two proteins^{18–20} (Fig. 4). The Cas6f protein has a two-domain architecture, which consists of an N-terminal ferredoxin-like fold similar to that in Cas6e, but its C-terminal domain is structurally distinct. An arginine-rich helix in the C-terminal domain of Cas6f inserts into the major groove of the crRNA duplex, and the bottom of the crRNA is positioned for sequence-specific hydrogen-bonding contacts in the RNA major groove. These contacts position the scissile phosphate of the crRNA in the enzyme active site so that cleavage occurs 8 nucleotides upstream of the spacer sequence^{18–20} (Fig. 4).

Although Cas6f and Cas6e recognize the sequence and shape of the crRNA hairpin in their respective systems, CRISPR RNA repeats in other CRISPR systems are thought to be unstructured³⁵. For example, the Cas6 protein from *P. furiosus* associates with CRISPR transcripts that are expected to contain unstructured repeats⁶⁴. The specific recognition of an unstructured RNA repeat requires a distinct mechanistic solution for RNA substrate discrimination. Remarkably, crystallographic studies of the Cas6 protein from *P. furiosus* have revealed the same duplicated ferredoxin-like fold observed in the Cas6e protein, but with a different mode of RNA recognition involving the opposite face of the protein (Fig. 4). In Cas6, the two ferredoxin-like folds clamp the 5' end of the single-stranded RNA repeat sequence in place²¹. Although the RNA in this structure is disordered in the enzyme active site, biochemical studies have shown that cleavage occurs 8 nucleotides upstream of the spacer sequence^{16,64}. While the nucleotide sequences at the cleavage site vary for each of the different Cas6 proteins, all Cas6 family endoribonucleases cleave their cognate RNA 8 nucleotides upstream of the spacer sequence using a metal-ion-independent mechanism.

Despite advances in our understanding of crRNA biogenesis, the diversity of *cas* genes has obscured identification of the protein factors responsible for CRISPR RNA processing in some systems. Type II immune systems consist of four *cas* genes, none of which have a detectable sequence similarity to known CRISPR-specific endoribonucleases. Recently, a different CRISPR RNA processing mechanism has been reported that involves RNase-III-mediated cleavage of double-stranded regions of the CRISPR RNA repeats¹⁷. The first indication of this mechanism came from deep sequencing of RNA from *S. pyogenes*. An abundant transcript containing a 25-nucleotide sequence that was complementary to the CRISPR repeat was identified. This RNA, termed tracrRNA (trans-activating CRISPR RNA), is coded on the opposite strand and just upstream of the CRISPR locus. Genetic and biochemical experiments demonstrated that tracrRNA and pre-crRNA are co-processed by RNase III, which produces cleavage products with a 2 nucleotide 3' overhang¹⁷. *In vivo* processing of CRISPR RNAs required Cas9 (previously known as Csn1), although a precise role for this enzyme in RNA processing has not yet been defined. The essential role of cellular proteins that are not solely involved in CRISPR-mediated defence, such as RNase III, indicates that different host factors may be involved as ancillary components of these immune systems.

crRNA-guided interference

The third stage of CRISPR-mediated immunity is target interference (Fig. 2). Here crRNAs associate with Cas proteins to form large CRISPR-associated ribonucleoprotein complexes that can recognize invading

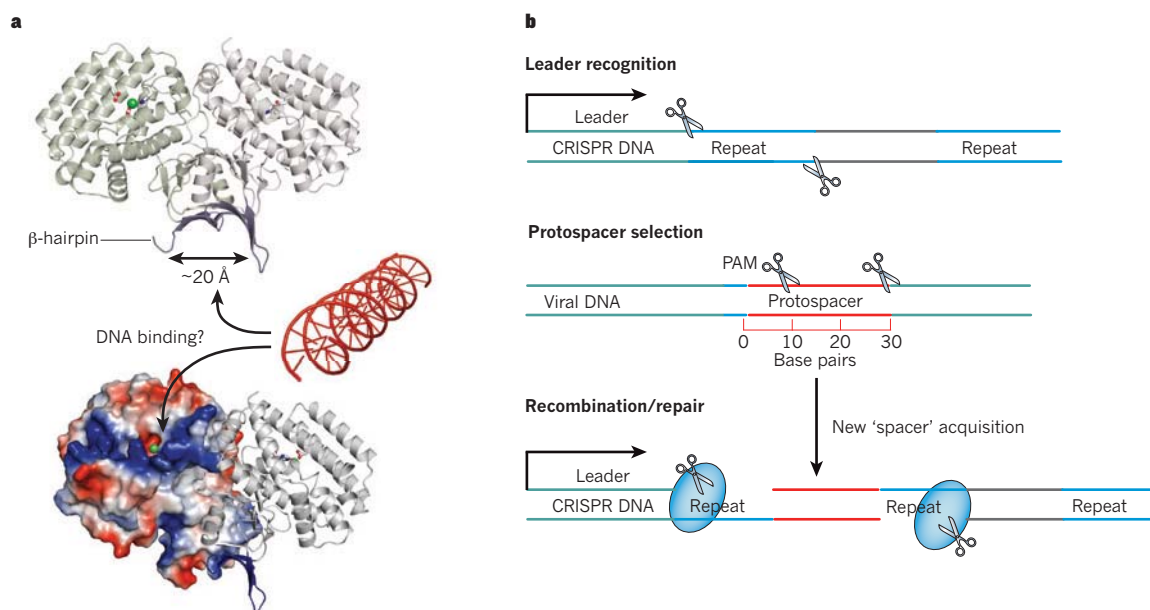


Figure 3 | Steps leading to new spacer integration. **a**, The Cas1 protein forms a stable homodimer where the two molecules (green and grey) are related by a pseudo-two-fold axis of symmetry (PBD ID: 3GOD)^{54,56}. This organization creates a saddle-like structure in the N-terminal domain, in which β-hairpins (blue) from each symmetrically related molecule hang (like stirrups) that are separated by approximately 20 Å, and may interact with the phosphodiester backbone of double-stranded DNA. An electrostatic surface representation (bottom) reveals a cluster of basic residues (blue) that form a positively charged strip across the metal-binding surface of the C-terminal domain. This strip may serve as an electrostatic trap that positions DNA substrates proximally to catalytic

metal ions (green sphere). **b**, CRISPR adaptation occurs by integrating fragments of foreign nucleic acid preferentially at the leader end of the CRISPR, forming new repeat-spacer units in the process. Protospacers are chosen non-randomly and may be selected from regions flanking the protospacer adjacent motif (PAM). Coordinated cleavage of the foreign DNA and integration of the protospacer into the leader-end of the CRISPR occurs through a mechanism that duplicates the repeat sequence and thus preserves the repeat-spacer-repeat architecture of the CRISPR locus. Although the protein components required for this process have not been conclusively identified, Cas1 and other general recombination or repair factors have been implicated (blue ovals)^{32,54,56}.

nucleic acids. Foreign nucleic acids are identified by base-pairing interactions between the crRNA spacer sequence and a complementary sequence from the intruder. Phage- and plasmid-challenge experiments performed in several model systems have demonstrated that crRNAs complementary to either the coding or the non-coding strand of the invading DNA can provide immunity^{14,22,47,60,65,66}. This is indicative of an RNA-guided DNA-targeting system, and indeed a pathway for DNA silencing has recently been demonstrated in *S. thermophilus*¹⁵. DNA sequencing and Southern blots indicated that both strands of the target DNA are cleaved within the region that is complementary to the crRNA spacer sequence¹⁵. This mechanism efficiently eliminates foreign DNA sequences, which have been specified by the spacer region of the crRNA, but avoids targeting the complementary DNA sequences in the CRISPR region of the host chromosome. The mechanism for distinguishing self from non-self is built into the crRNA. The spacer sequence of each crRNA is flanked by a portion of the adjacent CRISPR repeat sequence, and any complementarity beyond the spacer into the adjacent repeat region signals self and prevents the destruction of the host chromosome⁶⁷.

However, not all CRISPR systems target DNA. *In vitro* experiments using enzymes from the type III-B CRISPR system of *P. furiosus* have shown that this system cleaves target RNA rather than DNA⁴⁸. All DNA targeting systems encode a complementary DNA sequence for each crRNA in the CRISPR locus and therefore require a mechanism for distinguishing self (CRISPR locus) from non-self (invading DNA). In contrast, systems that target RNA may not be required to make this distinction because most CRISPR loci are transcribed only in one direction and thus do not generate complementary RNA targets. CRISPR systems that target RNA may be uniquely capable of defending against viruses that have RNA-based genomes. However, adaptation of the CRISPR in response to a challenge by an RNA-based virus will probably require the invading RNA to be reverse-transcribed into DNA before it can be integrated into the CRISPR locus.

Cas proteins directly participate in target binding. Recent biochemical studies have shown that CRISPR-associated complexes facilitate target recognition by enhancing sequence-specific hybridization between the CRISPR RNA and complementary target sequences²⁷. A short high-affinity binding site located at one end of the crRNA spacer sequence governs the efficiency of target binding, and viruses that acquired a single mismatch in this region were able to escape detection by the immune system⁶⁰. This high-affinity binding site is functionally analogous to the 'seed' sequence (Fig. 1) that has been identified in eukaryotic microRNAs (miRNAs)⁶⁸. Structural and biochemical studies have shown that Argonaute proteins facilitate target recognition by pre-ordering the nucleotides at the 5' end of the miRNA in a helical configuration⁶⁹. This pre-ordering reduces the entropic penalty that is associated with helix formation and provides a thermodynamic advantage for target binding within this region. A similar mechanism may occur during crRNA target binding, providing an interesting example of convergent evolution between CRISPR-based immunity in prokaryotes and RNAi in eukaryotes (Fig. 1).

Structural and biochemical studies have been performed on CRISPR-associated complexes isolated from three different type I CRISPR systems^{24–27,48}. These complexes seem to share some general morphological features, but the precise special arrangement of the Cas proteins and their interactions with the crRNA have been unclear. Sub-nanometre-resolution structures of the CRISPR-associated complex from *E. coli* (Cascade) have recently been determined using cryo-electron microscopy²⁶. This complex is comprised of an unequal stoichiometry of 5 functionally essential Cas proteins and a 61-nucleotide crRNA^{22,24,26}. The structure reveals a sea-horse-shaped architecture in which the crRNA is displayed along a helical arrangement of protein subunits that protect the crRNA from degradation²⁶. The 5' and 3' ends of the crRNA form unique structures

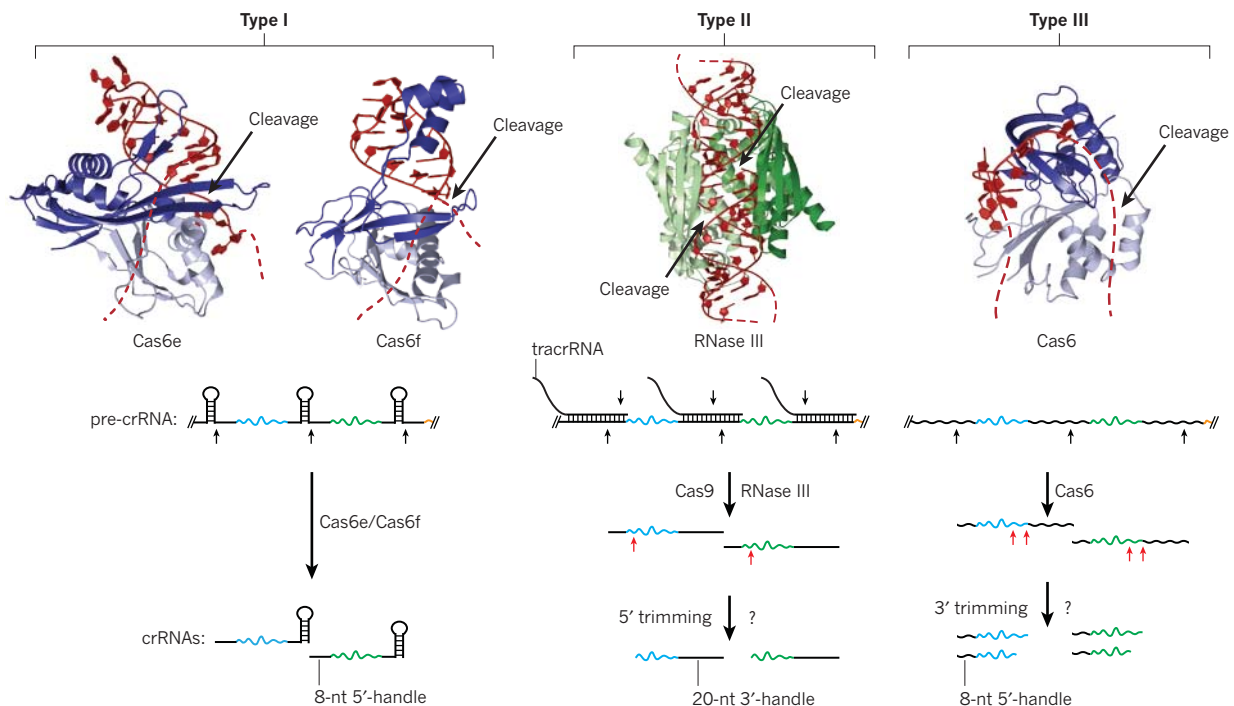


Figure 4 | Diverse mechanisms of CRISPR RNA biogenesis. CRISPR RNA repeats are specifically recognized and cleaved by diverse mechanisms. In type I CRISPR systems, Cas6e (PDB ID: 2Y8W) and Cas6f (PDB ID: 2XLK) recognize the major groove of the crRNA stem-loop primarily through electrostatic interactions using a β -hairpin and α -helix, respectively^{18,19,20}. Cleavage occurs at the double-stranded–single-stranded junction (black arrows), leaving an 8-nt 5'-handle on mature crRNAs. In type II CRISPR systems, tracrRNA hybridizes to the pre-crRNA repeat to form duplex RNAs that are substrates for endonucleolytic cleavage by host RNase III (PDB ID: 2EZ6), an activity that may also require Cas9 (ref. 17). Subsequent trimming (red arrows) by an unidentified nuclease

removes leftover repeat sequences from the 5' end. Cas6 (PDB ID: 3PKM) in type III-B CRISPR systems specifically recognizes single-stranded RNA, upstream of the scissile phosphate, on a face of the protein opposite that of the previously identified active site residues^{16,21,64}. The remainder of the repeat substrate probably wraps around the protein (red dashed line) to allow cleavage 8 nucleotides upstream of the repeat-spacer junction. Subsequent 3' trimming (red arrows) generates mature crRNAs of two discrete lengths. The N-terminal domain of all Cas 6 family proteins adopts a ferredoxin-like fold (light blue). The C-terminal domain of Cas6 and Cas6e also adopts a ferredoxin-like fold but the C-terminal domain of Cas6f is structurally distinct (dark blue).

that are anchored at opposite ends of the Cascade complex, displaying the 32-nucleotide spacer sequence for base-pairing with complementary targets.

The structure of Cascade bound to a 32-nucleotide target sequence²⁶ reveals a concerted conformational change that could be a signal for recruiting Cas3. Cas3 — the *trans*-acting nuclease of type I CRISPR systems — may function as a target 'slicer' in a similar way to Argonaute in RNAi pathways^{22,44,46,70}. Although Cas3 was implicated previously in the process of self versus non-self discrimination, recent studies have demonstrated that Cascade recognizes the PAM directly and that mutations in the PAM decrease Cascade's affinity for the target⁶⁰. The importance of the PAM is highlighted by the recovery of phage and plasmid escape mutants, which frequently contain a single mutation in the PAM^{15,51–53,60}. The structure of Cascade indicates that the PAM is positioned near the 'tail' of the sea-horse-shaped complex. High-resolution structures and mutational analysis of the nucleic acid and protein components in this and related systems are needed to determine the mechanisms of target authentication and degradation.

Applications of CRISPR structure and function

The sequence diversity of CRISPR loci, even within closely related strains, has been used for high-resolution genotyping and forensic medicine. This technique, known as spooligotyping (spacer oligotyping), has been applied successfully to the analysis of human pathogens, including *Mycobacterium tuberculosis*⁷¹, *Corynebacterium diphtheriae*⁷² and *Salmonella enterica*⁷³. Spooligotyping was developed long before the function of CRISPRs was understood, but now that studies have begun to reveal the biological function and mechanism of CRISPR-mediated genetic silencing, new opportunities for creative applications have emerged.

Laboratory strains of bacteria are grown in high-density bioreactors for many different applications in the food industry, and they are becoming increasingly important in the production of biofuels. CRISPR systems offer a natural mechanism for adapting economically important bacteria for resistance against multiple phages.

The biochemical activities of various Cas proteins may have useful applications in molecular biology in much the same way that DNA restriction enzymes have revolutionized cloning and DNA manipulation. A wide range of CRISPR-specific endoribonucleases that recognize small RNA motifs with high affinity expand the number of tools available for manipulating nucleic acids. In addition, a crRNA-guided ribonucleoprotein complex in *P. furiosus* was shown to cleave target RNAs⁴⁸. Site-specific cleavage of target RNA molecules could have a range of uses, from generating homogeneous termini after *in vitro* transcription to targeting a specific intracellular messenger RNA for inactivation in a similar way to RNAi. CRISPRs also provide a new mechanism for limiting the spread of antibiotic resistance or the transfer of virulence factors by blocking horizontal gene transfer^{15,47}. In addition, CRISPRs participate in a regulatory mechanism that alters biofilm formation in *P. aeruginosa*^{74,75}. Although the clinical relevance of CRISPRs remains to be demonstrated, the opportunities for creative implementation of this new gene-regulation system are perceptibly vast.

Future directions of CRISPR biology

The discovery of some of the fundamental mechanisms of CRISPR-based adaptive immunity has raised new questions and highlighted the areas with the greatest potential for future research. Although CRISPR RNA processing and targeting steps are now understood in some detail, how and when target sequences are identified during a phage infection

or plasmid transformation are still unclear. Furthermore, why DNA or RNA target sequences are chosen, and their fate once they are bound to a crRNA-targeting complex is not well understood. In addition, the mechanisms by which foreign sequences are selected and integrated into CRISPR loci are almost entirely unknown. Some CRISPR loci seem to be considerably more active than others, at least under laboratory conditions, so selection of the model organisms will be important. The diversity and prevalence of CRISPR systems throughout microbial communities ensures that new findings and applications in this field will be forthcoming in the years ahead. ■

1. Hoskisson, P. A. & Smith, M. C. Hypervariation and phase variation in the bacteriophage 'resistome'. *Curr. Opin. Microbiol.* **10**, 396–400 (2007).
2. Rodriguez-Valera, F. et al. Explaining microbial population genomics through phage predation. *Nature Rev. Microbiol.* **7**, 828–836 (2009).
3. Weinbauer, M. G. Ecology of prokaryotic viruses. *FEMS Microbiol. Rev.* **28**, 127–181 (2004).
4. Labrie, S. J., Samson, J. E. & Moineau, S. Bacteriophage resistance mechanisms. *Nature Rev. Microbiol.* **8**, 317–327 (2010).
5. Stern, A. & Sorek, R. The phage-host arms race: shaping the evolution of microbes. *Bioessays* **33**, 43–51 (2010).
6. Al-Attar, S., Westra, E. R., van der Oost, J. & Brouns, S. J. Clustered regularly interspaced short palindromic repeats (CRISPRs): the hallmark of an ingenious antiviral defense mechanism in prokaryotes. *Biol. Chem.* **392**, 277–289 (2011).
7. Deveau, H., Garneau, J. E. & Moineau, S. CRISPR/Cas system and its role in phage-bacteria interactions. *Annu. Rev. Microbiol.* **64**, 475–493 (2010).
8. Horvath, P. & Barrangou, R. CRISPR/Cas, the immune system of bacteria and archaea. *Science* **327**, 167–170 (2010).
9. Karginov, F. V. & Hannon, G. J. The CRISPR system: small RNA-guided defense in bacteria and archaea. *Mol. Cell* **37**, 7–19 (2010).
10. Makarova, K. S. et al. Evolution and classification of the CRISPR-Cas systems. *Nature Rev. Microbiol.* **9**, 467–477 (2011).
11. Marraffini, L. A. & Sontheimer, E. J. CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nature Rev. Genet.* **11**, 181–190 (2010).
12. Sorek, R., Kunin, V. & Hugenholtz, P. CRISPR—a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nature Rev. Microbiol.* **6**, 181–186 (2008).
13. Andersson, A. F. & Banfield, J. F. Virus population dynamics and acquired virus resistance in natural microbial communities. *Science* **320**, 1047–1050 (2008).
14. Barrangou, R. et al. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709–1712 (2007).
This study provided the first direct evidence for CRISPR immune system function, demonstrating that new spacer acquisition provides resistance against future phage infection in a cas gene-dependent manner.
15. Garneau, J. E. et al. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468**, 67–71 (2010).
This study showed that CRISPRs can acquire new spacers upon plasmid challenge and provided the first example of crRNA-guided cleavage of double-stranded DNA in vivo.
16. Carte, J., Wang, R., Li, H., Terns, R. M. & Terns, M. P. Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes Dev.* **22**, 3489–3496 (2008).
17. Deltcheva, E. et al. CRISPR RNA maturation by *trans*-encoded small RNA and host factor RNase III. *Nature* **471**, 602–607 (2011).
This study identified a new CRISPR RNA processing pathway that involves a *trans*-acting RNA complementary to the CRISPR repeat sequence and demonstrated that processing of this duplex is mediated by cellular RNase III.
18. Gesner, E. M., Schellenberg, M. J., Garside, E. L., George, M. M. & MacMillan, A. M. Recognition and maturation of effector RNAs in a CRISPR interference pathway. *Nature Struct. Mol. Biol.* **18**, 688–692 (2011).
19. Haurwitz, R. E., Jinek, M., Wiedenheft, B., Zhou, K. & Doudna, J. A. Sequence- and structure-specific RNA processing by a CRISPR endonuclease. *Science* **329**, 1355–1358 (2010).
20. Sashital, D. G., Jinek, M. & Doudna, J. A. An RNA induced conformational change required for CRISPR RNA cleavage by the endonuclease Cse3. *Nature Struct. Mol. Biol.* **18**, 680–687 (2011).
21. Wang, R., Preamplume, G., Terns, M. P., Terns, R. M. & Li, H. Interaction of the Cas6 ribonuclease with CRISPR RNAs: recognition and cleavage. *Structure* **19**, 257–264 (2011).
22. Brouns, S. J. et al. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**, 960–964 (2008).
This paper revealed that long CRISPR RNA transcripts are processed into small RNAs (crRNAs) by a dedicated endoribonuclease and that the processed RNAs are bound by a Cas protein complex (Cascade), which together with Cas3 are required for phage protection.
23. Hale, C., Kleppe, K., Terns, R. M. & Terns, M. P. Prokaryotic silencing (psi)RNAs in *Pyrococcus furiosus*. *RNA* **14**, 2572–2579 (2008).
24. Jore, M. M. et al. Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nature Struct. Mol. Biol.* **18**, 529–536 (2011).
25. Lintner, N. G. et al. Structural and functional characterization of an archaeal clustered regularly interspaced short palindromic repeat (crispr)-associated complex for antiviral defense (CASCADE). *J. Biol. Chem.* **286**, 21643–21656 (2011).
26. Wiedenheft, B. et al. Structures of the RNA-guided surveillance complex from a bacterial immune system. *Nature* **477**, 486–489 (2011).
This paper presented the first sub-nanometer resolution structures of Cascade, showing how the crRNA is accommodated within a large ribonucleoprotein complex that is involved in foreign nucleic acid surveillance.
27. Wiedenheft, B. et al. RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. *Proc. Natl Acad. Sci. USA* **108**, 10092–10097 (2011).
28. Ishino, Y., Shinagawa, H., Makino, K., Armemura, M. & Nakata, A. Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *J. Bacteriol.* **169**, 5429–5433 (1987).
29. Bolotin, A., Quinquis, B., Sorokin, A. & Ehrlich, S. D. Clustered regularly interspaced short palindromic repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151**, 2551–2561 (2005).
30. Mojica, F. J., Diez-Villasenor, C., Garcia-Martinez, J. & Soria, E. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol.* **60**, 174–182 (2005).
31. Pourcel, C., Salvignol, G. & Vergnaud, G. CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology* **151**, 653–663 (2005).
32. Makarova, K. S., Grishin, N. V., Shabalina, S. A., Wolf, Y. I. & Koonin, E. V. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol. Direct* **1**, 7 (2006).
33. Grissa, I., Vergnaud, G. & Pourcel, C. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res.* **35**, W52–W57 (2007).
34. Rousseau, C., Gonnet, M., Le Romancer, M. & Nicolas, J. CRISPI: a CRISPR interactive database. *Bioinformatics* **25**, 3317–3318 (2009).
35. Kunin, V., Sorek, R. & Hugenholtz, P. Evolutionary conservation of sequence and secondary structures in CRISPR repeats. *Genome Biol.* **8**, R61 (2007).
36. Bland, C. et al. CRISPR recognition tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. *BMC Bioinformatics* **8**, 209 (2007).
37. Dsouza, M., Larsen, N. & Overbeek, R. Searching for patterns in genomic data. *Trends Genet.* **13**, 497–498 (1997).
38. Edgar, R. C. PILER-CR: fast and accurate identification of CRISPR repeats. *BMC Bioinformatics* **8**, 18 (2007).
39. Jansen, R., Embden, J. D., Gaastra, W. & Schouls, L. M. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol. Microbiol.* **43**, 1565–1575 (2002).
40. Pougach, K. et al. Transcription, processing and function of CRISPR cassettes in *Escherichia coli*. *Mol. Microbiol.* **77**, 1367–1379 (2010).
41. Pul, U. et al. Identification and characterization of *E. coli* CRISPR-cas promoters and their silencing by H-NS. *Mol. Microbiol.* **75**, 1495–1512 (2010).
42. Westra, E. R. et al. H-NS-mediated repression of CRISPR-based immunity in *Escherichia coli* K12 can be relieved by the transcription activator LeuO. *Mol. Microbiol.* **77**, 1380–1393 (2010).
43. Haft, D. H., Selengut, J., Mongodin, E. F. & Nelson, K. E. A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS Comput. Biol.* **1**, e60 (2005).
44. Sinkunas, T. et al. Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system. *EMBO J.* **30**, 1335–1342 (2011).
45. Han, D. & Krauss, G. Characterization of the endonuclease SSO2001 from *Sulfolobus solfataricus* P2. *FEBS Lett.* **583**, 771–776 (2009).
46. Mulepati, S. & Bailey, S. Structural and biochemical analysis of the nuclease domain of the clustered regularly interspaced short palindromic repeat (CRISPR) associated protein 3 (CAS3). *J. Biol. Chem.* **286**, 31896–31903 (2011).
47. Marraffini, L. A. & Sontheimer, E. J. CRISPR interference limits horizontal gene transfer in *Staphylococci* by targeting DNA. *Science* **322**, 1843–1845 (2008).
48. Hale, C. R. et al. RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell* **139**, 945–956 (2009).
Most CRISPR systems appear to target DNA, but this article reported the isolation of a multisubunit ribonucleoprotein complex (Cmr-complex) from *P. furiosus* that cleaves target RNAs (not DNA) at a fixed distance from the 3' end of the crRNA-guide.
49. Tyson, G. W. & Banfield, J. F. Rapidly evolving CRISPRs implicated in acquired resistance of microorganisms to viruses. *Environ. Microbiol.* **10**, 200–207 (2008).
50. Snyder, J. C., Bateson, M. M., Lavin, M. & Young, M. J. Use of cellular CRISPR (clusters of regularly interspaced short palindromic repeats) spacer-based microarrays for detection of viruses in environmental samples. *Appl Environ Microbiol.* **76**, 7251–7258 (2010).
51. Deveau, H. et al. Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J. Bacteriol.* **190**, 1390–1400 (2008).
52. Horvath, P. et al. Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. *J. Bacteriol.* **190**, 1401–1412 (2008).
53. Sapranaukas, R. et al. The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*. *Nucleic Acids Res.* **39**, 9275–9282 (2011).
54. Babu, M. et al. A dual function of the CRISPR-Cas system in bacterial antiviral immunity and DNA repair. *Mol. Microbiol.* **79**, 484–502 (2011).
55. Han, D., Lehmann, K. & Krauss, G. SSO1450—a CAS1 protein from *Sulfolobus*

- solfataricus* P2 with high affinity for RNA and DNA. *FEBS Lett.* **583**, 1928–1932 (2009).
56. Wiedenheft, B. *et al.* Structural basis for DNase activity of a conserved protein implicated in CRISPR-mediated antiviral defense. *Structure* **17**, 904–912 (2009).
 57. Chen, C. S. *et al.* A proteome chip approach reveals new DNA damage recognition activities in *Escherichia coli*. *Nature Methods* **5**, 69–74 (2008).
 58. Aguilera, A. The connection between transcription and genomic instability. *EMBO J.* **21**, 195–201 (2002).
 59. Mojica, F. J., Diez-Villasenor, C., Garcia-Martinez, J. & Almendros, C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155**, 733–740 (2009).
 60. Semenova, E. *et al.* Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc. Natl Acad. Sci. USA* **108**, 10098–10103 (2011).
 61. Tang, T. H. *et al.* Identification of 86 candidates for small non-messenger RNAs from the archaeon *Archaeoglobus fulgidus*. *Proc. Natl Acad. Sci. USA* **99**, 7536–7541 (2002).
 62. Tang, T. H. *et al.* Identification of novel non-coding RNAs as potential antisense regulators in the archaeon *Sulfolobus solfataricus*. *Mol. Microbiol.* **55**, 469–481 (2005).
 63. Ebihara, A. *et al.* Crystal structure of hypothetical protein TTHB192 from *Thermus thermophilus* HB8 reveals a new protein family with an RNA recognition motif-like domain. *Protein Sci.* **15**, 1494–1499 (2006).
 64. Carte, J., Pfister, N. T., Compton, M. M., Terns, R. M. & Terns, M. P. Binding and cleavage of CRISPR RNA by Cas6. *RNA* **16**, 2181–2188 (2010).
 65. Gudbergsson, S. *et al.* Dynamic properties of the *Sulfolobus* CRISPR/Cas and CRISPR/Cmr systems when challenged with vector-borne viral and plasmid genes and protospacers. *Mol. Microbiol.* **79**, 35–49 (2011).
 66. Manica, A., Zebec, Z., Teichmann, D. & Schleper, C. *In vivo* activity of CRISPR-mediated virus defence in a hyperthermophilic archaeon. *Mol. Microbiol.* **80**, 481–491 (2011).
 67. Marraffini, L. A. & Sontheimer, E. J. Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature* **463**, 568–571 (2010).
- This study identified a mechanism for distinguishing self from non-self, which relies on base-pairing potential in regions outside the crRNA spacer sequence.**
68. Bartel, D. P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281–297 (2004).
 69. Parker, J. S., Parizotto, E. A., Wang, M., Roe, S. M. & Barford, D. Enhancement of the seed-target recognition step in RNA silencing by a PIWI/MID domain protein. *Mol. Cell* **33**, 204–214 (2009).
 70. Beloglazova, N. *et al.* Structure and activity of the Cas3 HD nuclease MJ0384, an effector enzyme of the CRISPR interference. *EMBO J.* **30**, 4616–4627 (2011).
 71. Groenen, P. M., Bunschoten, A. E., van Soolingen, D. & van Embden, J. D. Nature of DNA polymorphism in the direct repeat cluster of *Mycobacterium tuberculosis*; application for strain differentiation by a novel typing method. *Mol. Microbiol.* **10**, 1057–1065 (1993).
 72. Mokrousov, I., Limeschenko, E., Vyazovaya, A. & Narvskaya, O. *Corynebacterium diphtheriae* spoligotyping based on combined use of two CRISPR loci. *Biotechnol. J.* **2**, 901–906 (2007).
 73. Liu, F. *et al.* Novel virulence gene and clustered regularly interspaced short palindromic repeat (CRISPR) multilocus sequence typing scheme for subtyping of the major serovars of *Salmonella enterica* subsp. *enterica*. *Appl. Environ. Microbiol.* **77**, 1946–1956 (2011).
 74. Cady, K. C. & O'Toole, G. A. Non-identity targeting of yersinia-subtype CRISPR-prophage interaction requires the Csy and Cas3 proteins. *J. Bacteriol.* **193**, 3433–3445 (2011).
 75. Zegans, M. E. *et al.* Interaction between bacteriophage DMS3 and host CRISPR region inhibits group behaviors of *Pseudomonas aeruginosa*. *J. Bacteriol.* **191**, 210–219 (2009).
 76. Obbard, D. J., Gordon, K. H. J., Buck, A. H. & Jiggins, F. M. The evolution of RNAi as a defence against viruses and transposable elements. *Philos. Trans. R. Soc. B Biol. Sci.* **364**, 99–115 (2009).
 77. Aravin, A. A., Hannon, G. J. & Brennecke, J. The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. *Science* **318**, 761–764 (2007).
 78. Aravin, A. A. *et al.* Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the *D-melanogaster* germline. *Curr Biol* **11**, 1017–1027 (2001).
 79. Bartel, D. P. MicroRNAs: target recognition and regulatory functions. *Cell* **136**, 215–233 (2009).
 80. Guo, H., Ingolia, N. T., Weissman, J. S. & Bartel, D. P. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* **466**, 835–840 (2010).

Acknowledgements B.W. is a Howard Hughes Medical Institute Fellow of the Life Sciences Research Foundation. S.H.S. acknowledges support from the National Science Foundation and National Defense Science & Engineering Graduate Research Fellowship programs. J.A.D. is an Investigator of the Howard Hughes Medical Institute.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence should be addressed to J.A.D. (doudna@berkeley.edu).

Modular regulatory principles of large non-coding RNAs

Mitchell Guttman^{1,2} & John L. Rinn^{1,3}

It is clear that RNA has a diverse set of functions and is more than just a messenger between gene and protein. The mammalian genome is extensively transcribed, giving rise to thousands of non-coding transcripts. Whether all of these transcripts are functional is debated, but it is evident that there are many functional large non-coding RNAs (ncRNAs). Recent studies have begun to explore the functional diversity and mechanistic role of these large ncRNAs. Here we synthesize these studies to provide an emerging model whereby large ncRNAs might achieve regulatory specificity through modularity, assembling diverse combinations of proteins and possibly RNA and DNA interactions.

More than half a century after being placed as the central component in the flow of genetic information from gene to protein, it is now accepted that RNA can perform diverse roles. Shortly after the discovery of messenger RNA, a large class of heteronuclear RNAs (hnRNAs)¹ was described, which did not include mRNA or associate with polyribosomes². Following years of sifting through these hnRNAs, the first RNA subfamilies were identified. These included small nuclear RNAs involved in splicing regulation³ and small nucleolar RNAs involved in ribosome biogenesis⁴, as well as the ribosomal RNAs and transfer RNAs involved in translation^{5,6}.

The world of RNA genes became even more complex with the discovery of RNAs that resembled mRNA in length and splicing structure but did not code for proteins. The first example was H19, which was identified as an RNA that was induced during liver development in the mouse⁷. The mouse *H19* transcript contained no large open reading frames (ORFs), but instead only small sporadic ORFs that were not evolutionarily conserved, did not template translation *in vivo* and did not produce an identifiable protein product⁸. Shortly afterwards, another non-coding RNA (ncRNA), termed XIST, was found to be expressed exclusively from the inactive X chromosome⁹ and later demonstrated to be required for X inactivation in mammals¹⁰. Over the next two decades, more large ncRNA genes were discovered including *Airm*¹¹, *Tug1* (ref. 12), *NRON*¹³ and *HOTAIR*¹⁴. With the availability of a draft sequence of the human genome, it became clear that much of the mammalian genome is transcribed^{15–18}. These transcripts were mapped to discrete loci throughout the genome. Over the next 10 years, both large and small RNA transcripts were discovered at an unprecedented rate^{15,17–20}; however, the functional significance of most of these transcripts was unclear. Although some of these could be considered noise^{21,22}, there are still many large ncRNAs that are known to have diverse functions^{23–29}.

This Review focuses on the classic examples of large ncRNAs that have helped to form the basis of more recent global studies of coding potential, function and mechanism. We discuss the concepts that have emerged from these examples that provide a framework for understanding the principles of RNA interactions. We propose that by assembling distinct regulatory components, large ncRNAs could produce intricate functional specificity, which is suggestive of a possible modular RNA code.

RNA maps

After the sequencing of the human genome, the next major hurdle was to define the genes it encoded. To do this, several research groups developed tiling microarrays^{17,19,20} and complementary DNA sequencing

methods¹⁵ to investigate transcriptional activity across the human genome, which led to the observation of widespread transcription of the genome. These studies, although limited to specific tissues and cell types, demonstrated that the mammalian genome encodes many thousands of non-coding transcripts including both short (<200 nucleotides in length) and long (>200 nucleotides in length) transcripts. In this Review, we focus on large ncRNAs produced from long transcripts, including those that originate from intergenic loci or overlapping protein-coding genes.

Dramatic innovations in sequencing technologies have allowed the deep sequencing of cDNAs, known as RNA-Seq³⁰; this deep sequencing, coupled with new computational methods for assembling the transcriptome³¹, has identified non-coding transcripts across many different cell types and tissues^{31,32}. It is now clear that there are thousands of well-expressed large ncRNAs with exquisite cell-type and tissue specificity^{31–33}.

As the numbers of identified non-coding transcripts increased, so did the uncertainty regarding their function; this led some authors to express concern that many of these transcripts may be just transcriptional noise^{21,22} with no function or incidental by-products of transcription from enhancer regions^{34,35}. These concerns are supported by the observations that many of these transcripts are expressed at extremely low levels^{32,36} and they have lower levels of evolutionary conservation than protein-coding genes^{25,31,37}. Although some of these transcripts may indeed be transcriptional noise²¹, the remaining transcripts consist of many distinct subclasses, including processed small RNAs^{18,29,38}, promoter-associated RNAs^{18,39}, transcripts from enhancer regions^{34,35} and functional large ncRNAs^{14,23}; each class varies in its expression and conservation properties^{31,37}. Distinguishing between these classes of RNA transcripts requires additional biological information including the coding potential of the RNA and the chromatin modifications of the corresponding genomic region (Fig. 1a).

Chromatin signatures

Genomic DNA is wrapped around histone proteins and packaged into higher-order structures termed chromatin⁴⁰. These histones can be modified in different ways that are indicative of the underlying DNA functional state. Advances in sequencing technologies have allowed the comprehensive characterization of the chromatin-modification landscape of mammalian genomes^{41–44}. These studies revealed combinations of histone modifications (termed chromatin signatures) that correspond to various gene properties, including a signature for active

¹Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, Massachusetts 02142, USA. ²Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. ³Stem Cell and Regenerative Biology, Harvard University, Cambridge, Massachusetts 02138, USA.

transcription^{41,44}. This signature consists of a short stretch of trimethylation of histone protein H3 at the lysine in position 4 (H3K4me3), which corresponds to promoter regions, followed by a longer stretch of trimethylation of histone H3 at the lysine in position 36 (H3K36me3), which covers the entire transcribed region^{41,44} (Fig. 1a).

Chromatin maps revealed that, similar to protein-coding genes, many ncRNA genes also contain a 'K4–K36' signature⁴⁴. By searching

for K4–K36 domains that do not overlap with known genes, chromatin signatures revealed approximately 1,600 regions in the mouse genome and approximately 2,500 regions in the human genome that were actively transcribed^{25,45}. The vast majority of these intergenic K4–K36 domains produce multi-exonic RNAs that have little capability to encode a conserved protein^{25,31}. RNAs expressed from these K4–K36 domains were termed large intergenic ncRNAs (lincRNAs) because

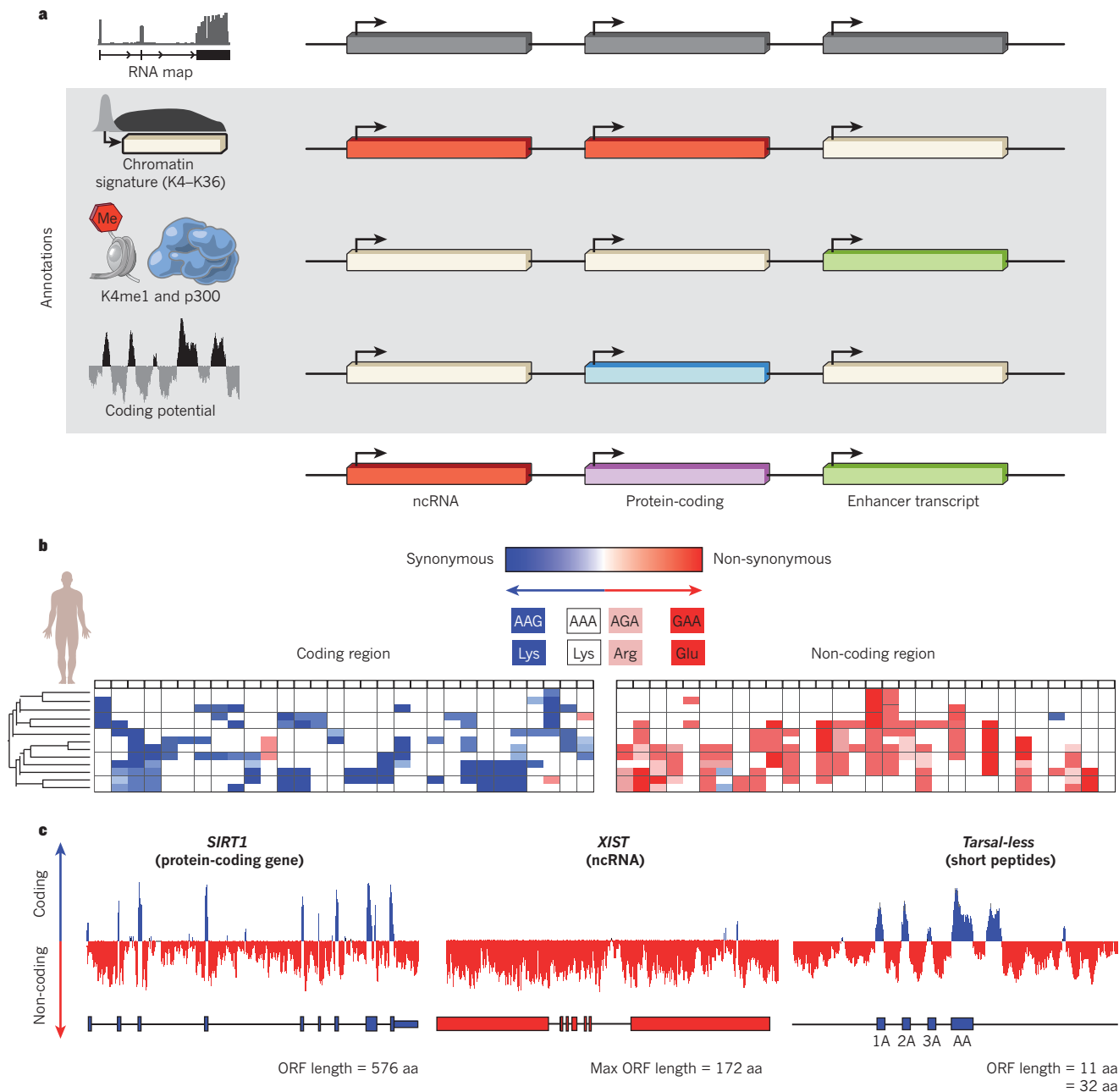


Figure 1 | Layering of genomic regions. **a**, Genomic regions are colour-coded by the presence of different genomic annotations. RNA transcription of a locus (grey), K4–K36 chromatin signature (red), K4me1 modification and transcriptional activator p300 (green) and protein-coding potential (blue). By overlaying this information, distinct transcripts are revealed, including ncRNAs (red), protein-coding genes (purple) and transcripts from enhancer regions (green). **b**, A cross-species alignment of a coding and a non-coding gene. Boxes represent codons, and each row represents a different aligned species. Blue boxes represent mutations that cause a synonymous substitution, and red boxes represent mutations that cause a non-synonymous substitution. A score capturing the coding potential of a sequence across species aligns sequences in all frames and scores mutations that maintain coding potential (blue boxes)

relative to mutations that break coding potential (that is, non-synonymous mutations, stop codons and frameshifting insertions or deletions) (red boxes). **c**, The coding potential score is shown for three gene types, *SIRT1* (a protein-coding gene), *XIST* (ncRNA gene) and *tarsal-less* (small-peptide coding gene), in which positive scores represent coding regions (blue) and negative scores represent non-coding regions (red). In each example, the gene structure is shown, where blue boxes represent known protein-coding exons and red boxes represent non-coding exons. *SIRT1* with an ORF length of 576 amino acids (aa) contains a positive score over each coding exon but not the non-coding regions. *XIST* with an ORF length of 172 amino acids contains negative scores over the entire transcribed region. *tarsal-less* with an ORF of 11 and 32 amino acids, contains positive scores over all known small peptides.

identification by this chromatin signature required the RNAs to be contained within the intergenic regions²⁵. Similarly, chromatin-state maps revealed that active enhancer regions contained short stretches of H3 lysine 4 monomethylation (H3K4me1) (ref. 43) and the transcriptional coactivator p300 (ref. 42), as well as additional modifications⁴⁶ (Fig. 1a). By coupling RNA sequencing and chromatin maps, many of the already identified non-coding transcripts were observed to be transcribed from active enhancers^{34,35}. However, lincRNAs and transcripts from enhancer regions are distinct classes, which are marked by different chromatin signatures^{25,34}. Although it needs to be determined whether transcripts originating from enhancers have a function^{34,35}, the functional importance of lincRNAs is becoming clearer^{14,23,24,26,28,47}. Several of these lincRNAs have been shown to have enhancer-like functions as they activate the expression of neighbouring genes^{24,28}.

Coding potential

Determining whether a transcript is non-coding is challenging because a long non-coding transcript is likely to contain an ORF purely by chance⁴⁸. Accordingly, the evidence for the absence of coding potential for the *XIST* and *H19* genes came from the lack of evolutionary conservation of the identified ORFs, the lack of homology to known protein domains and the inability to template significant protein production^{8,49}. These principles have been generalized to classify coding potential across thousands of transcripts by scoring conserved ORFs across dozens of species^{50,51}, by searching for homology in large protein-domain databases⁵², and by sequencing RNA associated with polyribosomes⁵³.

Computational methods such as the 'codon substitution frequency' algorithm^{50,51} leverage evolutionary information to determine whether an ORF is conserved across species and provide a general strategy for determining coding potential (Fig. 1b, c). Owing to the large number of available genome sequences, these methods have been used to accurately determine conserved coding potential in regions as small as 5 amino acids²⁵, which makes them extremely sensitive to the potentially small peptides, such as the 11 amino acid peptide encoded by the *tar-sal-less* gene^{54,55} (Fig. 1c). Despite their sensitivity, conservation-based methods may fail to detect newly evolved proteins because they do not contain a conserved ORF^{50,51}. However, because many ncRNAs show clear evolutionary constraint^{25,31,37} but no evolutionarily conserved ORF, this indicates that the observed evolutionary selection is not due to a newly evolved protein.

Experimental methods, such as ribosome profiling, have provided a strategy for identifying ribosome occupancy on RNA, which have been proposed as a method for distinguishing between coding and non-coding transcripts⁵³. However, this still needs to be tested because non-coding transcripts that show an association with the ribosome have not been shown to have a protein product^{53,56}. Importantly, an association of RNA with a ribosome alone cannot be taken as evidence of protein-coding potential because both the ncRNAs of *H19* and *TUG1* can be detected in the ribosome^{53,57} despite having clear roles as ncRNAs^{8,45,58,59}.

An alternative explanation for these observed associations is 'translational noise', spurious association that may lead to non-functional translation products²². Consistent with this, virtually all of the transcripts that have been suggested to encode small peptides by ribosome profiling⁵³ lack the evolutionary conservation of their proposed coding regions^{25,31}, which is in striking contrast to almost all known protein-coding genes⁶⁰, including the few well-characterized functional small peptides^{56,61,62} (Fig. 1c). Accordingly, identification of any new protein-coding gene requires the clear demonstration of the function of the protein product *in vivo*^{54,55}.

Global identification of ncRNA function

Identifying the functional role of an ncRNA requires direct perturbation experiments, such as loss-of-function and gain-of-function. Individual ncRNAs involved in specific processes have been functionally characterized (see ref. 63 for a review). For example, *XIST* is crucial for random

inactivation of the X chromosome¹⁰; Air is crucial for imprinting control at the *Igf2r* locus¹¹; HOTAIR affects expression of the *HOXD* gene family¹⁴, as well as other genes throughout the genome^{45,64,65}; HOTTIP affects expression of the *HOXA* gene family²⁸; lincRNA-RoR affects reprogramming efficiency⁴⁷; NRON affects NFAT transcription factor activity¹³; and Tug1 affects retina development through the regulation of the cell cycle¹². Although there are now many examples of large ncRNAs that are required for the correct regulation of gene expression, this is just one of many functions in which they are involved; ranging from telomere replication⁶⁶ to translation⁶⁷.

The global characterization of ncRNA function has proved to be challenging because, in most cases, it is unclear which phenotype to investigate¹³. One approach to classifying the putative function of ncRNAs uses 'guilt-by-association'²⁵. This approach associates ncRNAs with biological processes based on a common expression pattern across cell types and tissues (Fig. 2a) and can therefore identify groups of ncRNAs that are associated with specific cellular processes (Fig. 2b). This approach has been used to predict roles for hundreds of ncRNAs in diverse biological processes such as stem cell pluripotency, immune responses, neural processes and cell-cycle regulation^{25,27,36}.

Although these correlations cannot prove that ncRNAs have a function in these processes, they do provide a hypothesis for targeted loss-of-function experiments. For example, lincRNA-p21 was predicted to be associated with the p53-mediated DNA damage response²⁵, and indeed lincRNA-p21 was found to be a target of p53 and on perturbation was shown to regulate apoptosis in response to DNA damage²⁶. In the same way, the ncRNA PANDA (p21 associated ncRNA DNA damage activated) was implicated, and was demonstrated to have a function, in the regulation of apoptosis²⁷. Another ncRNA, lincEnc1 (ref. 25), was predicted to have a role in cell-cycle regulation in embryonic stem (ES) cells and has been shown in a separate study to affect the proliferation of ES cells⁶⁸.

Alternatively, global approaches can be used to determine function, such as systematic RNA interference (RNAi) knockdown followed by gene-expression profiling. Unlike correlation analysis, these perturbation-based experiments provide evidence for the function of an ncRNA²³. Methods to classify function using this approach are conceptually similar to guilt-by-association because the function can be inferred on the basis of the genes that are affected by loss of function of ncRNAs²³. A systematic perturbation study demonstrated that knockdown of the vast majority of lincRNAs expressed in ES cells had a major effect on gene expression²³. The gene-expression signatures revealed dozens of lincRNAs that block key lineage-commitment programs within ES cells and function in crucial ES cell regulatory and signalling pathways. Importantly, this study also identified 26 lincRNAs that are required to maintain the pluripotent state²³.

Not all non-coding transcripts are functional RNA molecules. Several examples of intergenic transcription have been identified in which the process of transcription alone changes the chromatin- and transcription-factor-binding landscape to allow activation and repression of neighbouring genes^{69,70}. Methods that degrade RNA after its transcription, such as RNAi, can distinguish between a functional RNA molecule and the process of transcription, on which there should be no observable effect after RNA degradation. Collectively, the genome-wide guilt-by-association approach and targeted and global perturbation studies have demonstrated that large ncRNAs have a crucial regulatory role in diverse biological processes^{23,25-27,32,47}.

cis- versus trans-regulatory mechanisms

The discovery that the *XIST* product was an ncRNA, led immediately to the suggestion of a model for how it could function in an allele-specific manner⁹. In theory, an ncRNA has an intrinsic *cis*-regulatory capacity because it can function while remaining tethered to its own locus^{9,71} (Fig. 2c), whereas an mRNA must be dissociated, exported and translated for it to function. Here we define a *cis*-regulator as one that exerts its function on a neighbouring gene on the same allele from which it is transcribed, and define a *trans*-regulator as one that

does not meet this criterion. Owing to the unique *cis*-regulatory capability of ncRNAs, it has been speculated that *cis*-regulation could be a common mechanism for large ncRNAs^{24,71}. However, global functional evidence strongly suggests that this is not the case (Box 1).

To distinguish *cis*- from *trans*-regulatory models, initial studies

have used correlation analysis and identified a significant correlation of expression between ncRNAs and their neighbouring protein-coding genes^{21,72}. However, several of these cases have been demonstrated to be *trans*-regulatory models, and the apparent correlations are due to shared upstream regulation (such as, lincRNA-p21 (ref. 26) and lincRNA-Sox2 (ref. 25)), positional correlation

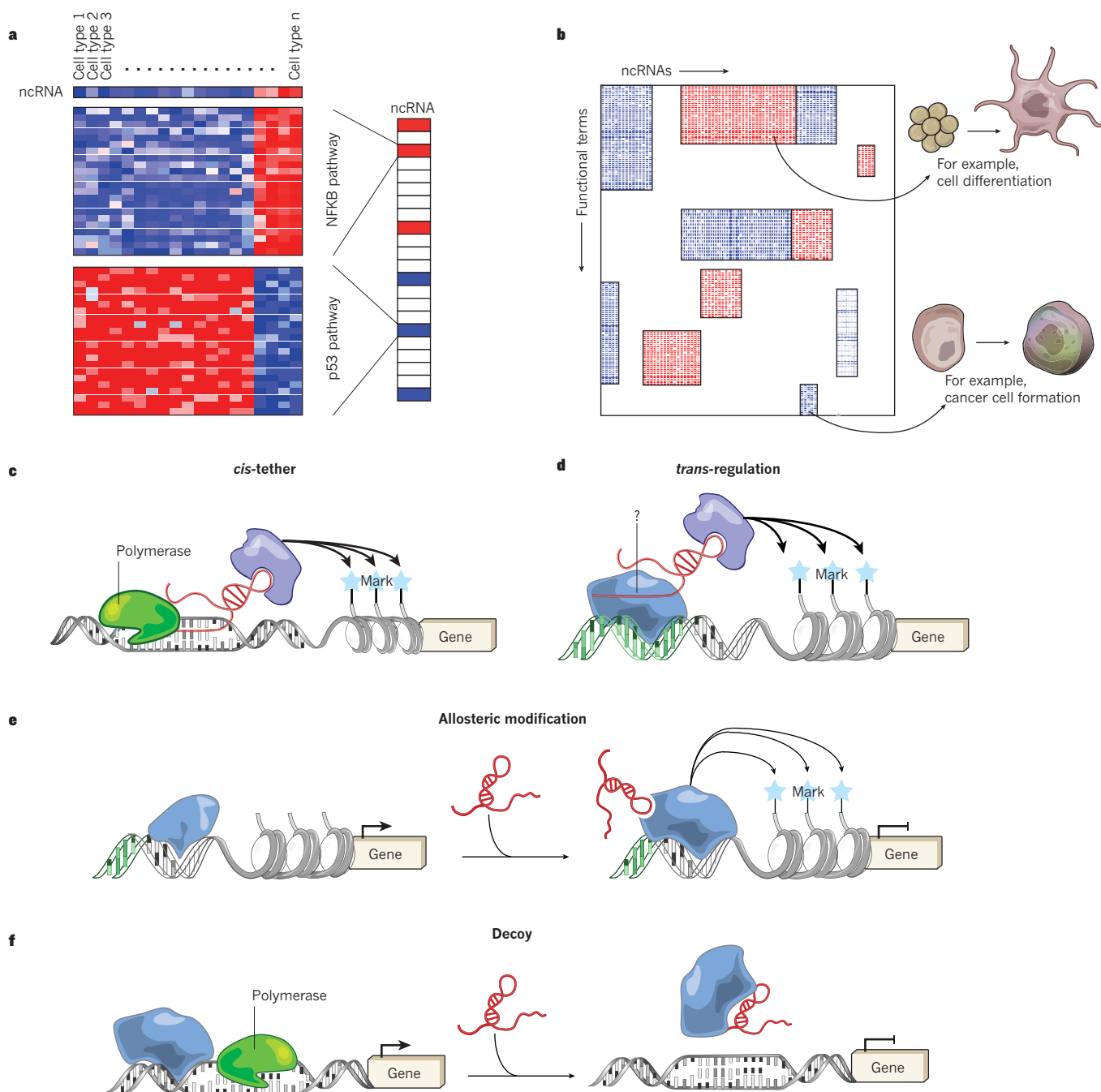


Figure 2 | Classification of ncRNA function. **a**, Illustration of an ncRNA with expression patterns related to the NFκB pathway. Each row represents a gene, and a positive association (red box) is assigned between the ncRNA and the pathway based on the correlation of the genes in the process. Similarly, the ncRNA is assigned negative association (blue box) with the p53 pathway based on anticorrelation with the genes in the process. **b**, The scores for each functional term and ncRNA can be clustered to identify classes of ncRNAs. In this example (adapted, with permission, from ref. 25) each column represents a different ncRNA, and each row represents a different functional term. **c**, A model of ncRNAs that have a *cis*-function by remaining tethered to their site of transcription. In this model, RNA

polymerase (green) transcribes an RNA (red), which can associate with regulatory proteins (purple) to affect neighbouring regions, as proposed for XIST^{9,71}. **d**, One model for ncRNA *trans*-regulation. In this model an ncRNA can associate with DNA-binding proteins (blue) and regulatory proteins to localize and affect the expression of the targets, as proposed for HOTAIR⁶⁴. **e**, A model for ncRNAs that bind regulatory proteins and change their activity, in this case leading to a change in modification state and expression of the target gene, as proposed for the CCND1 ncRNAs, which interact with the TLS protein⁸⁹. **f**, A model for ncRNAs that act as 'decoys'. In this model, ncRNAs bind protein complexes and prevent them from binding to their proper regulatory targets, as proposed for GAS5 and PANDA²⁷.

BOX 1

Distinguishing *cis*- from *trans*-regulation

If an ncRNA is a *cis*-regulator, then several observations will be true: (i) the gene-expression levels of a neighbouring gene will be correlated with the RNA expression across all conditions; (ii) loss-of-function of the RNA would affect expression of a neighbouring gene, and (iii) the ncRNA would affect expression of a neighbouring gene on the same allele that it is expressed from. The absence of any of these criteria supports *trans*-regulation. We illustrate this point using five common regulatory models. The figure shows what would be observed using specific computational and experimental methods for each regulatory model. The boxes with a tick indicate observed effects on neighbouring genes for each method, and boxes with a cross indicate no observed effect on neighbouring genes. Known ncRNA examples of each of these regulatory models are shown to the right of the figure.

	Regulatory model	Expression correlation	Perturbation effect	Allele-specific regulation	Known ncRNA examples
<i>trans</i>		×	×	×	
<i>trans</i>		✓	×	×	
<i>trans</i>		✓	✓	×	Unknown
<i>trans</i>		✓	✓	×	
<i>cis</i>		✓	✓	✓	
		✓ Neighbour affected	×	Neighbour unaffected	

(such as, HOTAIR¹⁴), transcriptional ‘ripple effects’²¹ and indirect regulation of neighbouring genes (Box 1). Consistent with these explanations, a recent study showed that an increased correlation of expression between ncRNAs and their neighbouring genes is comparable to that observed for protein-coding genes³².

Recently, loss-of-function experiments have been used to investigate *cis*- versus *trans*-effects of lincRNAs. One study knocked down seven lincRNAs and identified no effects on neighbouring genes but did show an effect on other genes⁴⁵. A second study knocked down 12 lincRNAs, 7 of which had modest effects on some of the genes within a wide genomic neighbourhood²⁴. More recently, a systematic study knocked down approximately 150 lincRNAs and identified no effect on the neighbouring genes for about 95% of the lincRNAs, which is similar to that observed for protein-coding genes²³.

Although perturbation experiments can demonstrate that an RNA functions as a *trans*-regulator, evidence for RNA acting as a *cis*-regulator is more difficult to obtain (Box 1). For example, perturbation experiments demonstrated that the ncRNA from JPX affects the expression of the neighbouring XIST gene, but as a *trans*-regulator⁷³. Conclusive proof of *cis*-regulation requires the demonstration that an RNA regulates a neighbouring gene on the same allele (Box 1). So far, few studies have performed this test, and it is unclear what percentage of ncRNAs that are suggested to have a *cis*-function by loss-of-function experiments^{24,28} will pass this test. Together, these studies indicate that although some ncRNAs are *cis*-regulators^{9,11,74–76}, the vast majority, which have been identified and characterized so far, function as *trans*-regulators^{14,23,26,45,73,77}.

Formation of RNA–protein interactions

The precise mechanism by which ncRNAs function remains poorly understood. However, one emerging theme is the interaction between ncRNAs and protein complexes. The functional importance of many ncRNA–protein interactions for correct transcriptional regulation has been demonstrated^{11,23,45,78–81}, including several ncRNAs that are required for the correct localization of chromatin proteins to genomic DNA targets^{79–83}.

The XIST ncRNA is a key example demonstrating that RNA can play a direct role in silencing large genomic regions⁸¹ by physically interacting with the polycomb complex⁸⁴, leading to the condensation of chromatin and transcriptional repression of an entire X chromosome⁸⁵ (Fig. 2c). Similar to XIST, many ncRNAs have been identified that physically associate with chromatin-regulatory complexes and ‘guide’ the associated complexes to specific genomic DNA regions, including HOTAIR¹⁴, AIR⁸⁶, KCNQ1ot1 (ref. 75) and lincRNA-p21 (ref. 26) (Fig. 2d).

Biochemical evidence has demonstrated that many large ncRNAs interact with chromatin regulators^{23,45,87,88}. The precise numbers vary depending on the experimental approach^{45,87}, but a conservative estimate suggests that at least 30% of lincRNAs associate with at least 1 of 12 distinct chromatin-regulatory complexes, which include readers, writers and erasers of chromatin modifications²³.

Importantly, lincRNAs can provide regulatory specificity to these complexes because the knockdown of these lincRNAs affects a subset of the genes that are normally regulated by these complexes^{23,45}. One hypothesis is that ncRNAs provide regulatory specificity by localizing chromatin-regulatory complexes to genomic DNA targets^{14,26,28,45,78,86}. Several methods have been developed to generate maps of RNA–DNA proximity^{82,83}, but it still needs to be determined what percentage of ncRNAs localize to genomic DNA regions and how these interactions occur.

In addition to their role in chromatin regulation, ncRNAs can also modulate the regulatory activity of protein complexes (Fig. 2e). As an example, an ncRNA upstream of cyclin D1 can bind to the TLS (translocation in liposarcoma) RNA-binding protein, which changes it from an inactive to an active state⁸⁹. Similarly, the NRON ncRNA can bind to the NFAT (nuclear factor of activated T cells)-transcription factor rendering it inactive because it prevents nuclear accumulation¹³. ncRNAs can also function as molecular ‘decoys’ by preventing correct regulation through competitive binding (Fig. 2f). For example, the GAS5 ncRNA binds to the glucocorticoid receptor and prevents the receptor from binding to its correct regulatory elements⁹⁰, and the PANDA ncRNA can prevent NF- κ B localization, which leads to apoptosis²⁷. Similarly, several studies have shown that ncRNAs can function as decoys to other RNA species, such as miRNAs, to control miRNA levels^{91,92}.

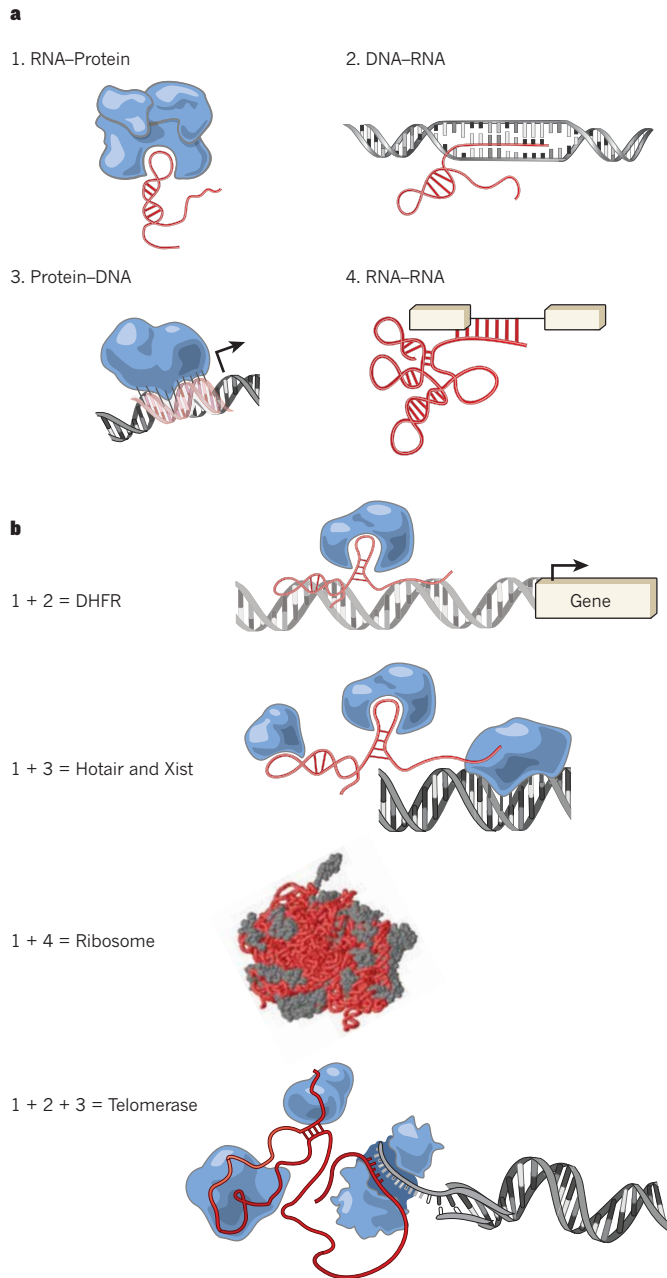


Figure 3 | Modular principles of large ncRNA genes. **a**, The four principles of nucleic acid and protein interactions. (1) RNA-protein interactions, (2) DNA-RNA hybridization-based interactions, (3) DNA-protein interactions and (4) RNA-RNA hybridization based interactions. **b**, Each of these principles can be combined to build distinct complexes. For example, combining RNA-protein and RNA-DNA interactions can localize a protein complex to a specific DNA sequence in an RNA-dependent manner; as has been implicated for the DHFR⁹⁹ promoter and localization of DNMT3b⁹⁸. Combining RNA-protein and protein-DNA principles can also localize a diverse set of proteins, which have a molecular scaffold created by RNA, to a specific DNA sequence in a protein-dependent manner. The ribosome is a multifaceted combination of RNA-protein interactions that facilitate correct RNA-RNA interactions for the ribozyme activity of the ribosome. The telomere replication activity of telomerase is an example of combining RNA-protein, RNA-DNA and protein-DNA interactions.

Large ncRNAs as molecular scaffolds of proteins

One emerging theme common to many large ncRNAs is the formation of multiple distinct RNA-protein interactions that are used to carry out their function (Fig. 3). The first indication of this phenomenon came from the discovery of telomerase⁹³. Telomerase activity requires

a telomerase RNA component (TERC)⁹⁴, which serves as a template for telomeric regulation and as a molecular scaffold for the polymerase enzyme around the RNA⁹⁵ (Fig. 3b). Importantly, genetic studies demonstrated that TERC plays a modular functional role, as genetically swapping particular domains of TERC retained the overall function⁶⁶. This indicated that TERC was made up of discrete functional modules to bring multiple proteins into the proximity of a protein⁶⁶.

More recently, HOTAIR was shown to contain distinct protein-interaction domains that can associate with polycomb repressive complex 2 (PRC2) (ref. 14) and the CoREST-LSD1 complex⁶⁴, which together are required for correct function (Fig. 3b). XIST also has discrete functional domains. Through a series of genetic deletions XIST was shown to contain at least two discrete domains that are responsible for silencing (RepA) and localization (RepC)⁸¹ (Fig. 3b). These functional domains could be independently deleted without affecting the role of the other domain, which suggests the modular nature of the XIST ncRNA⁸¹. These functional domains of XIST also interact with discrete proteins; the silencing domain (RepA) binds to PRC2 and the localization domain (RepC) binds to YY1 (ref. 96) and hnRNP⁹⁷. These examples show that large ncRNAs can function as molecular scaffolds of protein complexes. Importantly, this phenomenon is likely to be a general one because approximately 30% of ES cell lincRNAs associate with multiple regulatory complexes²³.

In addition to interacting with multiple proteins, in several examples, ncRNAs have been shown to interact directly with both DNA and RNA. ncRNAs for example form triplex structures with DNA^{98,99} (Fig. 3a) such as a ncRNA that binds to the ribosomal DNA promoter and interacts with the DNMT3b protein to silence expression⁹⁸. Furthermore, RNA can form traditional duplex base-pairing interactions with DNA, a property that has long been speculated for large ncRNAs⁷¹. Finally, RNA can form base-pair interactions with RNA (Fig. 3a), which are crucial for processes such as tRNA-mRNA anticodon recognition⁵, ribonuclease P recognition of pre-tRNAs⁵, miRNA targeting¹⁰⁰, ribosome structure as a ribozyme⁶⁷ and splicing regulation⁶. Despite these examples, the interactions between large ncRNAs, genomic DNA and other RNAs are not well characterized.

A potential modular RNA code

Collectively, the studies reviewed here suggest an intriguing hypothesis: large ncRNAs are flexible modular scaffolds^{23,64,66,81}. In this model, RNA contains discrete domains that interact with specific protein complexes. These RNAs, through a combination of domains, bring specific regulatory components into proximity with each other, which results in the formation of a unique functional complex. These RNA regulatory complexes can include interactions with proteins but can also extend to RNA-DNA and RNA-RNA regulatory interactions.

RNA is well-suited for this role because it is a malleable evolutionary substrate compared with a protein, allowing for the selection of discrete interaction domains⁵. Specifically, RNA can be easily mutated, tested and selected without breaking its core functionality⁵. This model of modular interactions can explain the observation that there are highly conserved 'patches' within large ncRNA genes^{25,31,37} that could have evolved for specific protein interactions^{26,81,84}. The remaining regions may be more evolutionarily flexible, allowing the formation of new functional domains by random mutation and selection. This is consistent with the observation that non-constrained regions of telomerase are dispensable⁶⁶.

The model of RNA as a modular scaffold is not limited to protein interactions. RNA can also base-pair with DNA, which might be used to guide complexes to specific DNA sequences. Alternatively, RNAs might guide complexes by bridging together sets of DNA-binding proteins. Such a model could explain how the same protein complexes are guided to different DNA loci in distinct cell types.

Large ncRNAs can also form RNA-RNA interactions, raising intriguing possibilities for future investigations. For example, two large RNA molecular scaffolds might be linked through RNA-RNA interactions. Another possibility is that RNA-RNA interactions could result in

unique RNA structures that can interact with protein complexes that are not attainable by the individual units. This has been observed in the ribosome, where the combination of RNA–RNA and RNA–protein interactions are required for correct complex formation.

Outlook

We are only beginning to understand the mechanism by which large ncRNAs carry out their regulatory function. A modular RNA regulatory code is an attractive hypothesis but remains to be tested; in particular, the way in which large ncRNAs, and proteins interact, and the underlying molecular principles are still unknown. Understanding these principles will require the identification of the sites of the RNA–protein interactions and the exact RNA-binding proteins *in vivo*. Furthermore, the way in which large ncRNAs localize to their target genes is unknown but could involve direct RNA–DNA interactions (Fig. 3a) or interactions with proteins that contain DNA recognition elements, which has been suggested for XIST⁹⁶ and HOTAIR⁶⁴. To gain insight into these processes, it will be important to catalogue the interactions that ncRNAs form with genomic DNA and RNAs. These data will help elucidate the rules that guide these interactions as well as the functional implications of these associations, which can then be tested experimentally.

If large ncRNAs are truly modular, then each individual domain would have a unique function that is independent of other domains. Demonstrating modularity will require the genetic deletion of domains and spacer regions, as well as domain-swapping experiments. Learning these principles would result in a defined ‘modular RNA code’ for how RNAs can affect cell states. By truly understanding this modular RNA code, it may be possible to create synthetically engineered RNAs that could interact with both nucleic acids and protein modules to carry out engineered regulatory roles. However, at present, it is premature to dismiss the possibility of large ncRNAs having other mechanisms of action that may not fit neatly into this modular RNA code. In the meantime, it is clear that mammalian genomes encode a diverse set of large important ncRNAs. ■

- Warner, J. R., Soeiro, R., Birnboim, H. C., Girard, M. & Darnell, J. E. Rapidly labeled HeLa cell nuclear RNA. I. Identification by zone sedimentation of a heterogeneous fraction separate from ribosomal precursor RNA. *J. Mol. Biol.* **19**, 349–361 (1966).
- Salditt-Georgieff, M., Harpold, M. M., Wilson, M. C. & Darnell, J. E., Jr. Large heterogeneous nuclear ribonucleic acid has three times as many 5' caps as polyadenylic acid segments, and most caps do not enter polyribosomes. *Mol. Cell. Biol.* **1**, 179–187 (1981).
- This paper demonstrates an abundant class of RNA species that do not enter polyribosomes.**
- Weinberg, R. A. & Penman, S. Small molecular weight monodisperse nuclear RNA. *J. Mol. Biol.* **38**, 289–304 (1968).
- Zieve, G. & Penman, S. Small RNA species of the HeLa cell: metabolism and subcellular localization. *Cell* **8**, 19–31 (1976).
- Gesteland, R. F., Cech, T. & Atkins, J. F. *The RNA World: The Nature of Modern RNA Suggests a Prebiotic RNA World*. 3rd edn (Cold Spring Harbor Laboratory Press, 2006).
- Eddy, S. R. Non-coding RNA genes and the modern RNA world. *Nature Rev. Genet.* **2**, 919–929 (2001).
- Pachnis, V., Brannan, C. I. & Tilghman, S. M. The structure and expression of a novel gene activated in early mouse embryogenesis. *EMBO J.* **7**, 673–681 (1988).
- Brannan, C. I., Dees, E. C., Ingram, R. S. & Tilghman, S. M. The product of the *H19* gene may function as an RNA. *Mol. Cell. Biol.* **10**, 28–36 (1990).
- This paper was the first report of a large ncRNA showing that the *H19* transcript lacked conserved ORFs and did not make a protein product *in vivo*.**
- Brown, C. J. *et al.* A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature* **349**, 38–44 (1991).
- Penny, G. D., Kay, G. F., Sheardown, S. A., Rastan, S. & Brockdorff, N. Requirement for *Xist* in X chromosome inactivation. *Nature* **379**, 131–137 (1996).
- Slutels, F., Zwart, R. & Barlow, D. P. The non-coding *Air* RNA is required for silencing autosomal imprinted genes. *Nature* **415**, 810–813 (2002).
- Young, T. L., Matsuda, T. & Cepko, C. L. The noncoding RNA taurine upregulated gene 1 is required for differentiation of the murine retina. *Curr. Biol.* **15**, 501–512 (2005).
- Willingham, A. T. *et al.* A strategy for probing the function of noncoding RNAs finds a repressor of NFAT. *Science* **309**, 1570–1573 (2005).
- Rinn, J. L. *et al.* Functional demarcation of active and silent chromatin domains in human *HOX* loci by noncoding RNAs. *Cell* **129**, 1311–1323 (2007).
- Carninci, P. *et al.* The transcriptional landscape of the mammalian genome. *Science* **309**, 1559–1563 (2005).
- This paper describes the large-scale cDNA sequencing efforts in the mouse genome and reveals many thousands of non-coding transcripts.**
- Birney, E. *et al.* Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799–816 (2007).
- Bertone, P. *et al.* Global identification of human transcribed sequences with genome tiling arrays. *Science* **306**, 2242–2246 (2004).
- Kapranov, P. *et al.* RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science* **316**, 1484–1488 (2007).
- Rinn, J. L. *et al.* The transcriptional activity of human Chromosome 22. *Genes Dev.* **17**, 529–540 (2003).
- Kapranov, P. *et al.* Large-scale transcriptional activity in chromosomes 21 and 22. *Science* **296**, 916–919 (2002).
- Ebisuya, M., Yamamoto, T., Nakajima, M. & Nishida, E. Ripples from neighbouring transcription. *Nature Cell Biol.* **10**, 1106–1113 (2008).
- Struhl, K. Transcriptional noise and the fidelity of initiation by RNA polymerase II. *Nature Struct. Mol. Biol.* **14**, 103–105 (2007).
- Guttman, M. *et al.* lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature* **477**, 295–300 (2011).
- Orom, U. A. *et al.* Long noncoding RNAs with enhancer-like function in human cells. *Cell* **143**, 46–58 (2010).
- Guttman, M. *et al.* Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* **458**, 223–227 (2009).
- This paper applied a chromatin signature to identify lincRNAs and used a guilt-by-association approach to classify their likely functions in diverse biological processes.**
- Huarte, M. *et al.* A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell* **142**, 409–419 (2010).
- Hung, T. *et al.* Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. *Nature Genet.* **43**, 621–629 (2011).
- Wang, K. C. *et al.* A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* **472**, 120–124 (2011).
- Wilusz, J. E., Freier, S. M. & Spector, D. L. 3' end processing of a long nuclear-retained noncoding RNA yields a tRNA-like cytoplasmic RNA. *Cell* **135**, 919–932 (2008).
- Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods* **5**, 621–628 (2008).
- Guttman, M. *et al.* *Ab initio* reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. *Nature Biotechnol.* **28**, 503–510 (2010).
- Cabili, M. N. *et al.* Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* **25**, 1915–1927 (2011).
- Mercer, T. R., Dinger, M. E., Sunkin, S. M., Mehler, M. F. & Mattick, J. S. Specific expression of long noncoding RNAs in the mouse brain. *Proc. Natl Acad. Sci. USA* **105**, 716–721 (2008).
- De Santa, F. *et al.* A large fraction of extragenic RNA Pol II transcription sites overlap enhancers. *PLoS Biol.* **8**, e1000384 (2010).
- Kim, T. K. *et al.* Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**, 182–187 (2010).
- Ravasi, T. *et al.* Experimental validation of the regulated expression of large numbers of non-coding RNAs from the mouse genome. *Genome Res.* **16**, 11–19 (2006).
- Ponjavic, J., Ponting, C. P. & Lunter, G. Functionality or transcriptional noise? Evidence for selection within long noncoding RNAs. *Genome Res.* **17**, 556–565 (2007).
- Taft, R. J. *et al.* Tiny RNAs associated with transcription start sites in animals. *Nature Genet.* **41**, 572–578 (2009).
- Seila, A. C. *et al.* Divergent transcription from active promoters. *Science* **322**, 1849–1851 (2008).
- Kouzarides, T. Chromatin modifications and their function. *Cell* **128**, 693–705 (2007).
- Barski, A. *et al.* High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823–837 (2007).
- Visel, A. *et al.* ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457**, 854–858 (2009).
- Heintzman, N. D. *et al.* Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**, 108–112 (2009).
- Mikkelsen, T. S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553–560 (2007).
- Khalil, A. M. *et al.* Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc. Natl Acad. Sci. USA* **106**, 11667–11672 (2009).
- Ernst, J. *et al.* Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**, 43–49 (2011).
- Loewer, S. *et al.* Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. *Nature Genet.* **42**, 1113–1117 (2010).
- Dinger, M. E., Pang, K. C., Mercer, T. R. & Mattick, J. S. Differentiating protein-coding and noncoding RNA: challenges and ambiguities. *PLoS Comput. Biol.* **4**, e1000176 (2008).
- Brockdorff, N. *et al.* The product of the mouse *Xist* gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. *Cell* **71**, 515–526 (1992).
- Lin, M. F., Deoras, A. N., Rasmussen, M. D. & Kellis, M. Performance and scalability of discriminative metrics for comparative gene identification in

- 12 *Drosophila* genomes. *PLoS Comput. Biol.* **4**, e1000067 (2008).
51. Lin, M. F., Jungreis, I. & Kellis, M. PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions. *Bioinformatics* **27**, i275–i282 (2011).
 52. Finn, R. D. *et al.* The Pfam protein families database. *Nucleic Acids Res.* **38**, D211–D222 (2010).
 53. Ingolia, N. T., Lareau, L. F. & Weissman, J. S. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* **147**, 789–802 (2011).
 54. Galindo, M. I., Pueyo, J. I., Fouix, S., Bishop, S. A. & Couso, J. P. Peptides encoded by short ORFs control development and define a new eukaryotic gene family. *PLoS Biol.* **5**, e106 (2007).
This paper demonstrates the existence of functional small peptides within a presumed 'non-coding' transcript through ORF conservation, *in vivo* protein identification and functional analysis.
 55. Kondo, T. *et al.* Small peptides switch the transcriptional activity of Shavenbaby during *Drosophila* embryogenesis. *Science* **329**, 336–339 (2010).
 56. Jiao, Y. & Meyerowitz, E. M. Cell-type specific analysis of translating RNAs in developing flowers reveals new levels of control. *Mol. Syst. Biol.* **6**, 419 (2010).
 57. Li, Y. M. *et al.* The *H19* transcript is associated with polysomes and may regulate *IGF2* expression in *trans*. *J. Biol. Chem.* **273**, 28247–28252 (1998).
 58. Cai, X. & Cullen, B. R. The imprinted *H19* noncoding RNA is a primary microRNA precursor. *RNA* **13**, 313–316 (2007).
 59. Yang, L. *et al.* ncRNA- and Pc2 methylation-dependent gene relocation between nuclear structures mediates gene activation programs. *Cell* **147**, 773–788 (2011).
 60. Clamp, M. *et al.* Distinguishing protein-coding and noncoding genes in the human genome. *Proc. Natl Acad. Sci. USA* **104**, 19428–19433 (2007).
 61. Kastenmayer, J. P. *et al.* Functional genomics of genes with small open reading frames (sORFs) in *S. cerevisiae*. *Genome Res.* **16**, 365–373 (2006).
 62. Hanada, K., Zhang, X., Borevitz, J. O., Li, W. H. & Shiu, S. H. A large number of novel coding small open reading frames in the intergenic regions of the *Arabidopsis thaliana* genome are transcribed and/or under purifying selection. *Genome Res.* **17**, 632–640 (2007).
 63. Mattick, J. S. The genetic signatures of noncoding RNAs. *PLoS Genet.* **5**, e1000459 (2009).
 64. Tsai, M. C. *et al.* Long noncoding RNA as modular scaffold of histone modification complexes. *Science* **329**, 689–693 (2010).
This paper identified multiple protein-interaction domains within HOTAIR that together allowed it to carry out its function, which demonstrated that a large ncRNA can act as a molecular scaffold.
 65. Gupta, R. A. *et al.* Long non-coding RNA *HOTAIR* reprograms chromatin state to promote cancer metastasis. *Nature* **464**, 1071–1076 (2010).
 66. Zappulla, D. C. & Cech, T. R. Yeast telomerase RNA: a flexible scaffold for protein subunits. *Proc. Natl Acad. Sci. USA* **101**, 10024–10029 (2004).
This paper demonstrated that telomerase RNA can bridge proteins by showing that protein interaction domains can be swapped and spacer regions deleted with minimal impact on the function of the RNA.
 67. Korostelev, A. & Noller, H. F. The ribosome in focus: new structures bring new insights. *Trends Biochem. Sci.* **32**, 434–441 (2007).
 68. Ivanova, N. *et al.* Dissecting self-renewal in stem cells with RNA interference. *Nature* **442**, 533–538 (2006).
 69. Martens, J. A., Laprade, L. & Winston, F. Intergenic transcription is required to repress the *Saccharomyces cerevisiae* *SER3* gene. *Nature* **429**, 571–574 (2004).
 70. Schmitt, S., Prestel, M. & Paro, R. Intergenic transcription through a Polycomb group response element counteracts silencing. *Genes Dev.* **19**, 697–708 (2005).
 71. Lee, J. T. Lessons from X-chromosome inactivation: long ncRNA as guides and tethers to the epigenome. *Genes Dev.* **23**, 1831–1842 (2009).
 72. Ponjavic, J., Oliver, P. L., Lunter, G. & Ponting, C. P. Genomic and transcriptional co-localization of protein-coding and long non-coding RNA pairs in the developing brain. *PLoS Genet.* **5**, e1000617 (2009).
 73. Tian, D., Sun, S. & Lee, J. T. The long noncoding RNA, *Jpx*, is a molecular switch for X chromosome inactivation. *Cell* **143**, 390–403 (2010).
 74. Koerner, M. V., Pauler, F. M., Huang, R. & Barlow, D. P. The function of non-coding RNAs in genomic imprinting. *Development* **136**, 1771–1783 (2009).
 75. Pandey, R. R. *et al.* *Kcnq1ot1* antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Mol. Cell* **32**, 232–246 (2008).
 76. Bertani, S., Sauer, S., Bolotin, E. & Sauer, F. The noncoding RNA *Mistral* activates *Hoxa6* and *Hoxa7* expression and stem cell differentiation by recruiting MLL1 to chromatin. *Mol. Cell* **43**, 1040–1046 (2011).
 77. Feng, J. *et al.* The *Evf-2* noncoding RNA is transcribed from the *Dlx-5/6* ultraconserved region and functions as a *Dlx-2* transcriptional coactivator. *Genes Dev.* **20**, 1470–1484 (2006).
 78. Koziol, M. J. & Rinn, J. L. RNA traffic control of chromatin complexes. *Curr. Opin. Genet. Dev.* **20**, 142–148 (2010).
 79. Maison, C. *et al.* Higher-order structure in pericentric heterochromatin involves a distinct pattern of histone modification and an RNA component. *Nature Genet.* **30**, 329–334 (2002).
 80. Bernstein, E. *et al.* Mouse polycomb proteins bind differentially to methylated histone H3 and RNA and are enriched in facultative heterochromatin. *Mol. Cell Biol.* **26**, 2560–2569 (2006).
 81. Wutz, A., Rasmussen, T. P. & Jaenisch, R. Chromosomal silencing and localization are mediated by different domains of *Xist* RNA. *Nature Genet.* **30**, 167–174 (2002).
This paper reported the generation of deletion mutants across the *Xist* locus and identified the discrete domains responsible for the silencing and localization roles of the RNA.
 82. Chu, C., Qu, K., Zhong, F. L., Artandi, S. E. & Chang, H. Y. Genomic maps of long noncoding RNA occupancy reveal principles of RNA–chromatin interactions. *Mol. Cell* **44**, 667–678 (2011).
 83. Simon, M. D. *et al.* The genomic binding-sites of a non-coding RNA. *Proc. Natl Acad. Sci. USA* **108**, 20497–20502 (2011).
 84. Zhao, J., Sun, B. K., Erwin, J. A., Song, J. J. & Lee, J. T. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* **322**, 750–756 (2008).
 85. Plath, K., Mlynarczyk-Evans, S., Nusinow, D. A. & Panning, B. *Xist* RNA and the mechanism of X chromosome inactivation. *Annu. Rev. Genet.* **36**, 233–278 (2002).
 86. Nagano, T. *et al.* The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science* **322**, 1717–1720 (2008).
 87. Zhao, J. *et al.* Genome-wide identification of Polycomb-associated RNAs by RIP-seq. *Mol. Cell* **40**, 939–953 (2010).
 88. Kaneko, S. *et al.* Phosphorylation of the PRC2 component Ezh2 is cell cycle-regulated and up-regulates its binding to ncRNA. *Genes Dev.* **24**, 2615–2620 (2010).
 89. Wang, X. *et al.* Induced ncRNAs allosterically modify RNA-binding proteins in *cis* to inhibit transcription. *Nature* **454**, 126–130 (2008).
 90. Kino, T., Hurt, D. E., Ichijo, T., Nader, N. & Chrousos, G. P. Noncoding RNA Gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor. *Sci. Signal* **3**, ra8 (2010).
 91. Salmena, L., Poliseno, L., Tay, Y., Kats, L. & Pandolfi, P. P. A ceRNA hypothesis: the Rosetta stone of a hidden RNA language? *Cell* **146**, 353–358 (2011).
 92. Cesana, M. *et al.* A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. *Cell* **147**, 358–369 (2011).
 93. Greider, C. W. & Blackburn, E. H. Identification of a specific telomere terminal transferase activity in *Tetrahymena* extracts. *Cell* **43**, 405–413 (1985).
 94. Feng, J. *et al.* The RNA component of human telomerase. *Science* **269**, 1236–1241 (1995).
 95. Lingner, J. *et al.* Reverse transcriptase motifs in the catalytic subunit of telomerase. *Science* **276**, 561–567 (1997).
 96. Jeon, Y. & Lee, J. T. YY1 tethers *Xist* RNA to the inactive X nucleation center. *Cell* **146**, 119–133 (2011).
 97. Hasegawa, Y., Brockdorff, N., Kawano, S., Tsutui, K. & Nakagawa, S. The matrix protein hnRNP U is required for chromosomal localization of *Xist* RNA. *Dev. Cell* **19**, 469–476 (2010).
 98. Schmitz, K. M., Mayer, C., Postepska, A. & Grummt, I. Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes. *Genes Dev.* **24**, 2264–2269 (2010).
 99. Martianov, I., Ramadass, A., Serra Barros, A., Chow, N. & Akoulitchiev, A. Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. *Nature* **445**, 666–670 (2007).
 100. Bartel, D. P. MicroRNAs: target recognition and regulatory functions. *Cell* **136**, 215–233 (2009).

Acknowledgements We thank M. Cabili, J. Engreitz, M. Garber, P. McDonel and A. Pauli for their reading and suggestions; T. Cech for comments and suggestions; E. Lander for helpful discussions and ideas; and S. Knemeyer and L. Gaffney for assistance with figures in this Review.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence should be addressed to M.G. (mguttman@mit.edu) and J.L.R. (john_rinn@harvard.edu).

The microcosmos of cancer

Amaia Lujambio¹ & Scott W. Lowe^{1,2}

The discovery of microRNAs (miRNAs) almost two decades ago established a new paradigm of gene regulation. During the past ten years these tiny non-coding RNAs have been linked to virtually all known physiological and pathological processes, including cancer. In the same way as certain key protein-coding genes, miRNAs can be deregulated in cancer, in which they can function as a group to mark differentiation states or individually as bona fide oncogenes or tumour suppressors. Importantly, miRNA biology can be harnessed experimentally to investigate cancer phenotypes or used therapeutically as a target for drugs or as the drug itself.

MicroRNAs (miRNAs) are small, evolutionarily conserved, non-coding RNAs of 18–25 nucleotides in length that have an important function in gene regulation. Mature miRNA products are generated from a longer primary miRNA (pri-miRNA) transcript through sequential processing by the ribonucleases Drosha and Dicer1 (ref. 1). The first description of miRNAs was made in 1993 in *Caenorhabditis elegans* as regulators of developmental timing^{2,3}. Later, miRNAs were shown to inhibit their target genes through sequences that are complementary to the target messenger RNA, leading to decreased expression of the target protein¹ (Box 1). This discovery resulted in a pattern shift in our understanding of gene regulation because miRNAs are now known to repress thousands of target genes and coordinate normal processes, including cellular proliferation, differentiation and apoptosis. The aberrant expression or alteration of miRNAs also contributes to a range of human pathologies, including cancer.

The control of gene expression by miRNAs is a process seen in virtually all cancer cells. These cells show alterations in their miRNA expression profiles, and emerging data indicate that these patterns could be useful in improving the classification of cancers and predicting their behaviour. In addition, miRNAs have now been shown to behave as cancer ‘drivers’ in the same way as protein-coding genes whose alterations actively and profoundly contribute to malignant transformation and cancer progression. Owing to the capacity of miRNAs to modulate tens to hundreds of target genes, they are emerging as important factors in the control of the ‘hallmarks’ of cancer⁴. In this Review, we summarize the findings that provide evidence for the central role of miRNAs in controlling cellular transformation and tumour progression. We also highlight the potential uses of miRNAs and miRNA-based drugs in cancer therapy and discuss the obstacles that will need to be overcome.

miRNAs are cancer genes

In 2002, Croce and colleagues first demonstrated that an miRNA cluster was frequently deleted or downregulated in chronic lymphocytic leukaemia⁵. This discovery suggested that non-coding genes were contributing to the development of cancer, and paved the way for the closer investigation of miRNA loss or amplification in tumours. Subsequently, miRNAs were shown to be differentially expressed in cancer cells, in which they formed distinct and unique miRNA expression patterns⁶, and whole classes of miRNAs could be controlled directly by key oncogenic transcription factors⁷. In parallel, studies with mouse models established that miRNAs were actively involved in tumorigenesis⁸. Collectively, these findings provided the first key insights into the relevance of miRNA biology in human cancer.

Despite these results, the sheer extent of involvement of miRNAs in cancer was not anticipated. miRNA genes are usually located in small chromosomal alterations in tumours (in amplifications, deletions or linked to regions of loss of heterozygosity) or in common chromosomal breakpoints that are associated with the development of cancer⁹. In addition to structural genetic alterations, miRNAs can also be silenced by promoter DNA methylation and loss of histone acetylation¹⁰. Interestingly, somatic translocations in miRNA target sites can also occur, representing a drastic means of altering miRNA function^{11,12}. The frequent deregulation of individual or clusters of miRNAs at multiple levels mirrors the deregulation for protein-coding oncogenes or tumour suppressors (Table 1).

In principle, somatic mutations that change an miRNA seed sequence could lead to the aberrant repression of tumour-suppressive mRNAs, but these seem to be infrequent¹³. Further sequencing could change this view, but this observation suggests that the intensity of miRNA signalling (altered by miRNA overexpression or underexpression) is more crucial than the specificity of the response. However, recent data indicate that miRNAs with an altered sequence can be produced through variable cleavage sites for Drosha and Dicer1, and that the presence of these variants can be perturbed in cancer¹⁴. Although the function of the variant ‘isomiRs’ remains unclear, in principle they could alter the quality of miRNA effects. State-of-the-art sequencing techniques will help to unmask mutations or modifications that otherwise would remain undetected. Whatever the mechanism, the widespread alteration in the expression of miRNAs is a ubiquitous feature of cancer.

miRNAs as cancer classifiers

Aberrant miRNA levels reflect the physiological state of cancer cells and can be detected by miRNA expression profiling and harnessed for the purpose of diagnosis and prognosis^{15,16}. In fact, miRNA profiling can be more accurate at classifying tumours than mRNA profiling because miRNA expression correlates closely with tumour origin and stage, and can be used to classify poorly differentiated tumours that are difficult to identify using a standard histological approach^{6,17}. Whether or not this increased classification power relates to the biology of miRNAs or the reduced complexity of the miRNA genome still needs to be determined.

The special features of miRNAs make them potentially useful for detection in clinical specimens. For example, miRNAs are relatively resistant to ribonuclease degradation, and they can be easily extracted from small biopsies, frozen samples and even formalin-fixed, paraffin-embedded tissues¹⁸. Furthermore, relatively simple and reproducible assays have been developed to detect the abundance of individual

¹Cancer Biology and Genetics Program, Memorial Sloan-Kettering Cancer Center (MSKCC), 1275 York Avenue, New York, New York 10065, USA. ²Howard Hughes Medical Institute, MSKCC, 1275 York Avenue, New York, New York 10065, USA.

BOX 1

Biogenesis and function of miRNAs

miRNAs are subjected to a unique biogenesis that is closely related to their regulatory functions. As the pathway in Fig. 1 shows, in general miRNAs are transcribed by RNA polymerase II into primary transcripts called pri-miRNAs⁷⁶. The primary transcripts contain a 5' cap structure a poly(A)⁺ tail and may include introns, similar to the transcripts of protein-coding genes⁷⁶. They also contain a region in which the sequences are not perfectly complementary, known as the stem-loop structure, which is recognized in the nucleus by the ribonuclease Drosha and its partner DGCR8, giving rise to the precursor miRNA (pre-miRNA) by cropping⁷⁶. However, some intronic miRNAs (called mirtrons) bypass the Drosha processing step and, instead, use splicing machinery to generate the pre-miRNA⁹⁹. The pre-miRNA is exported from the nucleus to the cytoplasm by XPO5 and is further cleaved by the ribonuclease Dicer1 (along with TARBP2) into a double-stranded miRNA (process known as dicing)⁷⁶. Again, this cleavage can be substituted by Argonaute-2-mediated processing¹⁰⁰.

After strand separation, the guide strand or mature miRNA forms, in combination with Argonaute proteins, the RNA-induced silencing complex (RISC), whereas the passenger strand is usually degraded. The mature strand is important for specific-target mRNA recognition and its consequent incorporation into the RISC¹. The specificity of miRNA targeting is defined by how complementary the 'seed' sequence (positions 2 to 8 from the 5' end of the miRNA) and the 'seed-match' sequence (generally in the 3' untranslated region of the target mRNA) are. The expression of the target mRNAs is silenced by miRNAs, either by mRNA cleavage ('slicing') or by translational repression¹. In addition, miRNAs have a number of unexpected functions, including the targeting of DNA, ribonucleoproteins or increasing the expression of a target mRNA⁹³. Overall, data indicate the complexity of miRNA-mediated gene regulation and highlight the importance of a better understanding of miRNA biology.

miRNAs, and methods that combine small RNA isolation, PCR and next-generation sequencing, allow accurate and quantitative assessment of all the miRNAs that are expressed in a patient specimen, including material that has been isolated by laser capture microdissection. The detection of global miRNA expression patterns for the diagnosis of cancers has not yet been proved; however, some individual or small groups of miRNAs have shown promise. For example, in non-small cell lung cancer, the combination of high miR-155 and low let-7 expression correlates with a poor prognosis, and in chronic lymphocytic leukaemia a 13 miRNA signature is associated with disease progression^{15,16}. Further advances in the technology of miRNA profiling could help to revolutionize molecular pathology.

Perhaps the most appealing application of miRNAs as a cancer diagnostic tool comes from the discovery of circulating miRNAs in serum. For example, miR-141 expression levels in serum were significantly higher in patients with prostate cancer than in healthy control individuals¹⁹. Although the analysis of circulating miRNAs is only just beginning, the successful advancement of this technology could provide a relatively non-invasive diagnostic tool for single-point or longitudinal studies. With such diagnostic tools in place, miRNA profiling could be used to guide cancer classification, facilitate treatment decisions, monitor treatment efficacy and predict clinical outcome.

When miRNA biogenesis goes awry

Although the expression of some miRNAs is increased in malignant cells, the widespread underexpression of miRNAs is a more common phenomenon. Whether this tendency is a reflection of a pattern associated with specific cells of origin, is a consequence of the malignant state or actively contributes to cancer development is still unclear. Because miRNA expression generally increases as cells differentiate, the apparent underexpression of miRNAs in cancer cells may, in part, be a result of miRNAs being 'locked' in a less-differentiated state. Alternatively, changes in oncogenic transcription factors that repress miRNAs or variability in the expression or activity of the miRNA processing machinery could also be important.

Two main mechanisms have been proposed as the underlying cause of the global downregulation of miRNAs in cancer cells. One involves transcriptional repression by oncogenic transcription factors. For example, the MYC oncoprotein, which is overexpressed in many cancers, transcriptionally represses certain miRNAs, although the extent to which this mediates its oncogenic activity or reflects a peripheral effect is still unknown²⁰. The other mechanism proposed involves changes in miRNA biogenesis and is based on the observation that cancer cells often display reduced levels, or altered activity, of factors in the miRNA biogenesis pathway²¹ (Box 1, Fig. 1).

In vivo studies have provided the most direct evidence of an active role for miRNA downregulation in at least some types of cancer. For example, analysis of mouse models in which the core enzymes of miRNA biogenesis have been constitutively or conditionally disrupted by different mechanisms suggests that these molecules function as haploinsufficient tumour suppressors. Thus, the repression of miRNA processing by the partial depletion of Dicer1 and Drosha accelerates cellular transformation and tumorigenesis *in vivo*²². Furthermore, deletion of a single *Dicer1* allele in lung epithelia promotes *Kras*-driven lung adenocarcinomas, whereas complete ablation of *Dicer1* causes lethality because of the need for miRNAs in essential processes²³. Consistent with the potential relevance of these mechanisms, reduced Dicer1 and Drosha levels have been associated with poor prognosis in the clinic²⁴. In addition to the core machinery, modulators of miRNA processing can also function as haploinsufficient tumour suppressors. Hence, point mutations that affect *TARBP2* or *XPO5* are correlated with sporadic and hereditary carcinomas that have microsatellite instability^{25,26}. Other miRNA modulators that influence the processing of only a subset of miRNAs could also be important. For example, LIN28A and LIN28B can bind and repress members of the *let-7* family (which are established tumour-suppressor miRNAs; Table 1), but this binding can be counteracted by KHSRP (KH-type splicing regulatory protein), also a factor involved in miRNA biogenesis; together this binding and counteracting dictate the level of mature let-7. The processing of miRNAs can be regulated by other genes including *DDX5* (helicase p68) or the SMAD 1 and SMAD 5 proteins, which may contribute to cancer development through the deregulation of miRNAs²⁷. Collectively, the global changes in miRNA expression that are seen in cancer cells probably arise through multiple mechanisms; the combined small changes in the expression of many miRNAs seem to have a large impact on the malignant state.

miRNAs as cancer drivers

Functional studies show that miRNAs that are affected by somatic alterations in tumours can affect cancer phenotypes directly, therefore confirming their driver function in malignancy. As drivers of malignancy, mechanistic studies show that these miRNAs interact with known cancer networks; hence, tumour-suppressor miRNAs can negatively regulate protein-coding oncogenes, whereas oncogenic miRNAs often repress known tumour suppressors (Fig. 2a). Perhaps the best example of this is the oncogenic miR-17-92 cluster, in which individual miRNAs suppress negative regulators of phosphatidylinositol-3-OH kinase signalling or pro-apoptotic members of the BCL-2 family, which disrupts the processes that are known to influence cancer development²⁸ (Table 1).

Cancer-associated miRNAs can also alter the epigenetic landscape

Table 1 | Key microRNAs involved in cancer

MicroRNA	Function	Genomic location	Mechanism	Targets	Cancer type	Mouse models	Clinical application
miR-17-92 cluster	Oncogene	13q22	Amplification and transcriptional activation	BIM, PTEN, CDKN1A and PRKAA1	Lymphoma, lung, breast, stomach, colon and pancreatic cancer	Cooperates with MYC to produce lymphoma. Overexpression induces lymphoproliferative disease	Inhibition and detection
miR-155	Oncogene	21q21	Transcriptional activation	SHIP1 and CEBPB	Chronic lymphocytic leukaemia, lymphoma, lung, breast and colon cancer	Overexpression induces pre-B-cell lymphoma and leukaemia	Inhibition and detection
miR-21	Oncogene	17q23	Transcriptional activation	PTEN, PDCD4 and TPM1	Chronic lymphocytic leukaemia, acute myeloid leukaemia, glioblastoma, pancreatic, breast, lung, prostate, colon and stomach cancer	Overexpression induces lymphoma	Inhibition and detection
miR-15a/16-1	Tumour suppressor	13q31	Deletion, mutation and transcriptional repression	BCL2 and MCL1	Chronic lymphocytic leukaemia, prostate cancer and pituitary adenomas	Deletion causes chronic lymphocytic leukaemia	Expression with mimics and viral vectors
let-7 family	Tumour suppressor	11 copies (multiple locations)	Transcriptional repression	KRAS, MYC and HMGA2	Lung, colon, stomach, ovarian and breast cancer	Overexpression suppresses lung cancer	Expression with mimics and viral vectors
miR-34 family	Tumour suppressor	1p36 and 11q23	Epigenetic silencing, transcriptional repression and deletion	CDK4, CDK6, MYC and MET	Colon, lung, breast, kidney, bladder cancer, neuroblastoma and melanoma	No published studies	Expression with mimics and viral vectors
miR-29 family	Oncogene	7q32 and 1q30	Transcriptional activation	ZFP36	Breast cancer and indolent chronic lymphocytic leukaemia	Overexpression induces chronic lymphocytic leukaemia	No published studies
	Tumour suppressor		Deletion and transcriptional repression	DNMTs	Acute myeloid leukaemia, aggressive chronic lymphocytic leukaemia and lung cancer		No published studies

BCL2, B-cell lymphoma protein-2; BIM, BCL-2-interacting mediator of cell death; CDKN1A, cyclin-dependent kinase inhibitor 1A; CEBPB, CCAAT/enhancer binding protein β ; HMGA2, high mobility group AT-hook 2; CDK4, cyclin-dependent kinase 4; CDK6, cyclin-dependent kinase 6; DNMT, DNA methyltransferase; MCL1, myeloid cell leukaemia sequence 1; PTEN, phosphatase and tensin homologue; PRKAA1, protein kinase, AMP-activated, alpha 1 catalytic subunit; PDCD4, programmed cell death 4; SHIP1, Src homology 2 domain-containing inositol 5-phosphatase 1; TPM1, tropomyosin 1; ZFP36, zinc finger protein 36.

of cancer cells. The cancer 'epigenome' is characterized by global and gene-specific changes in DNA methylation, histone modification patterns and chromatin-modifying enzyme expression profiles, which impact gene expression in a heritable way²⁹. In one way, miRNA expression can be altered by DNA methylation or histone modifications in cancer cells^{10,30}, but miRNAs can also regulate components of the epigenetic machinery, therefore indirectly contributing to the reprogramming of cancer cells. For example, miR-29 inhibits *DNMT3A* and *DNMT3B* expression in lung cancer³¹, whereas miR-101 regulates the histone methyltransferase *EZH2* in prostate cancer³². The presence of mature miRNAs in the nucleus³³ is another indication of the potentially direct role that miRNAs have in controlling epigenetic modifications, such as DNA methylation and histone modifications — a hypothesis that has been established in plants³⁴ but still needs to be demonstrated with certainty in mammals.

In the same way as protein-coding genes, miRNAs can be oncogenes or tumour suppressors depending on the cellular context in which they are expressed, which means that defining their precise contribution to cancer can be a challenge (Fig. 2b). The fact that miRNAs show tissue-specific expression and their output, shown in the cell's physiology, is dependent on the expression pattern of the specific mRNAs that harbour target sites could explain this apparent paradox. For example, the miR-29 family has a tumour-suppressive effect in lung tumours but appears oncogenic in breast cancer because of its ability to target the DNA methyltransferases *DNMT3A* and *DNMT3B*, and *ZFP36*, respectively^{31,35} (Table 1).

To further complicate the process, some miRNAs repress several positive components of a pathway, whereas others target both positive and negative regulators, possibly to buffer against minor physiological variations that could trigger much larger changes in the cell physiology³⁶. In cancer cells, this buffering role can mean that some miRNAs could

simultaneously target oncogenes and tumour-suppressor genes. In addition, combinations of miRNAs can cooperate to regulate one or several pathways, which increases the flexibility of regulation but confounds experimentalists³⁷ (Fig. 2c). Consequently, the way in which miRNAs contribute to cancer development is conceptually similar to cancer-associated transcription factors such as MYC and p53, which are mediated through many targets that depend on contextual factors that are influenced by cell type and micro-environment. From a practical perspective it is crucial that miRNA targets are studied in a context that is appropriate to the environment that is being studied to determine what impact they will have on tumour cell behaviour (Fig. 2b).

Oncogenic pathways

Beyond the impact of somatic genetic and epigenetic lesions, the altered expression of miRNAs in cancer can arise through the aberrant activity of transcription factors that control their expression. Interestingly, the same transcription factors are often targets of miRNA-mediated repression, which gives rise to complex regulatory circuits and feedback mechanisms. Thus, a single transcription factor can activate or repress several miRNAs and protein-coding genes; in turn, the alteration in miRNA expression can affect more protein-coding genes that then amplifies the effects of a single gene.

As already mentioned, MYC directly contributes to the global transcriptional silencing of miRNAs²⁰. This repression involves the downregulation of miRNAs with antiproliferative, antitumorigenic and pro-apoptotic activity such as, let-7, miR-15a/16-1, miR-26a or miR-34 family members³⁸ (Fig. 2d; Table 1). Initial studies indicate that Myc uses both transcriptional and post-transcriptional mechanisms to modulate miRNA expression. This phenomenon could be due to LIN28A and LIN28B being the direct target of MYC, and that they are required for

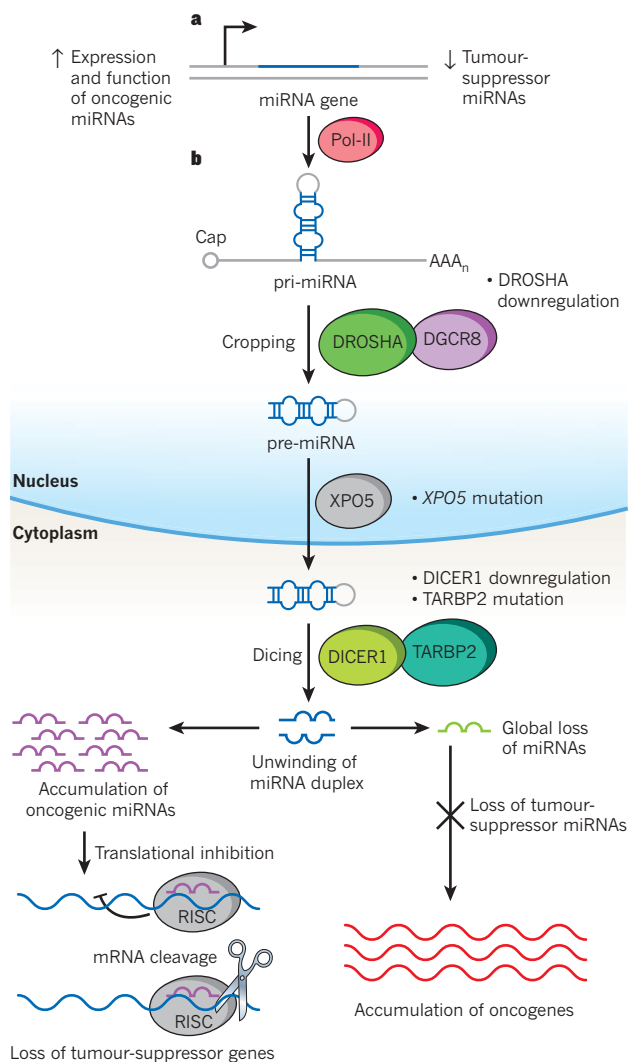


Figure 1 | Mechanisms of miRNA perturbation in cancer. Cancer cells present global downregulation of miRNAs, loss of tumour-suppressor miRNAs and specific accumulation of oncogenic miRNAs. The alteration in miRNA expression patterns leads to the accumulation of oncogenes and downregulation of tumour-suppressor genes, which leads to the promotion of cancer development. **a**, The expression and function of oncogenic miRNAs is increased by genomic amplification, activating mutations, loss of epigenetic silencing and transcriptional activation. By contrast, tumour-suppressor miRNAs are lost by genomic deletion, inactivating mutations, epigenetic silencing or transcriptional repression. **b**, After transcription, global levels of miRNAs can be reduced by impaired miRNA biogenesis. Inactivating mutations and reduced expression have been described for almost all the members of the miRNA processing machinery. If there is a downregulation of DROSHA this can lead to a decrease in the cropping of primary miRNA (pri-miRNA) to precursor miRNA (pre-miRNA). In the case of *XPO5* mutation, pre-miRNAs are prevented from being exported to the cytoplasm. Mutation of *TARBP2* or downregulation of *DICER1* results in a decrease in mature miRNA levels. Pol II, RNA polymerase II; RISC, RNA-induced silencing complex.

MYC-mediated repression of let-7 (ref. 38). Furthermore, MYC directly activates the transcription of miR-17-92 polycistronic cluster and, given its oncogenic role, it may contribute to MYC-induced tumorigenesis³⁹. MYC-driven reprogramming of miRNA expression could also be a factor in hepatocellular carcinoma, because of the contribution the reprogramming has to the aggressive phenotype of tumours originating from hepatic progenitor cells⁴⁰. Some miRNAs, such as let-7, also regulate MYC, closing the regulatory circuit³⁷.

miRNAs are embedded in many other oncogenic networks, including *KRAS* activation, which leads to the repression of several miRNAs. For

example, in pancreatic cancer with mutant *KRAS*, RAS-responsive element-binding protein 1 (*RREB1*) represses miR-143 and miR-145 promoter, and at the same time both *KRAS* and *RREB1* are targets of miR-143 and miR-145, revealing a feedforward mechanism that increases the effect of RAS signalling⁴¹. Similarly, *KRAS* is a target for several miRNAs, of which the let-7 family is the most representative example⁴². The integration of miRNAs into key oncogenic pathways, and the generation of feedforward and feedback loops that have a balancing effect, creates intricate ways to incorporate intracellular and extracellular signals in the decisions of cell proliferation or survival, and further implicates miRNAs in the pathogenesis of cancer.

TP53 is a master regulator of miRNAs

The *TP53* tumour suppressor is perhaps the most important and well-studied cancer gene, and it is not surprising that several studies have suggested that miRNA biology can have a role in its regulation and activity (Fig. 2e). The p53 protein acts as a sequence-specific DNA-binding factor that can activate and repress transcription. Although there is no doubt that most of the actions of p53 can be explained by its ability to control canonical protein-coding targets such as *CDKN1A* and *PUMA*, it can also transactivate several miRNAs. One of the best-studied classes is the miR-34 family (Table 1), which represses genes that can promote proliferation and apoptosis — plausible targets in a p53-mediated tumour-suppressor response⁴³. In principle, the action of p53 to induce the expression of miR-34 and other miRNAs can explain some of its transcriptional repressive functions.

The discovery of additional p53-regulated miRNAs, and the targeting of p53 or its pathway by other miRNAs, has provided general insights into the miRNA-mediated control of gene expression and the potential therapeutic opportunities for targeting the p53 network (Fig. 2e). Several p53-activated miRNAs, such as miR-192, miR-194, miR-215 and miR-605, can target *MDM2*, which is a negative regulator of p53 and a therapeutic target. These potentially relevant miRNAs can be epigenetically silenced in some types of cancer; however, their reactivation or reintroduction (see the section miRNAs as drugs and drug targets) offers an intriguing therapeutic opportunity for inhibiting *MDM2* in tumours that harbour wild-type p53 (refs 44, 45). Similarly, p53 can also activate miR-107, miR-200 or miR-192, which are miRNAs that inhibit angiogenesis and epithelial-to-mesenchymal transition^{46–48}. Conversely, p53 can be repressed by certain oncogenic miRNAs including miR-380-5p, which is upregulated in neuroblastomas with *MYCN* amplification, or miR-504, which decreases p53-mediated apoptosis and cell-cycle arrest and can promote tumorigenesis^{49,50}. However, the extent to which these miRNAs control life and death decisions in the p53 network still needs to be shown decisively to determine whether these miRNAs are valid therapeutic targets.

The studies mentioned have extended our understanding of the roles and regulation of p53 into the world of small non-coding RNAs, but the action on miRNA biology may be even more complex. For example, one study⁵¹ suggests that p53 can affect miRNA biogenesis by promoting pri-miRNA processing through association with the large Drosha complex (Fig. 2e), but the precise mechanism remains unclear⁵¹. In a more conventional way, the p53 family member p63 transcriptionally controls *Dicer1* expression. Mutant *TP53* can interfere with this regulation, which leads to a reduction in *Dicer1* levels and reduces the levels of certain cancer-relevant miRNAs⁵². Thus, with the p53 network as a typical example, it is clear that miRNAs can interact with cancer-relevant pathways at multiple and unexpected levels and that a better understanding of miRNA biology will help to decipher the role and function of other important cancer genes.

Micromanagement of metastasis and beyond

In addition to promoting cancer initiation, miRNAs can modulate processes that support cancer progression, including metastasis^{53–56}. As indicated earlier, changes in miRNA levels can occur through effects on their transcription or by global changes in the RNA interference (RNAi)

machinery, and both mechanisms seem to be important for this process. For example, in breast cancer, miR-10b and miR-9 can induce metastasis, whereas miR-126, miR-335 and miR-31 act as suppressors. The miR-200 family inhibits epithelial-to-mesenchymal transition, which influences one aspect of the metastatic process⁵⁷. However, miR-200 could also promote the colonization of metastatic cells in breast cancer, which provides yet another example of the opposing activities of some miRNAs⁵⁸. Conversely, in head and neck squamous-cell carcinomas, lung adenocarcinomas and breast cancers, the reduced levels of certain miRNAs that arise from *Dicer1* downregulation also promote cell motility and are associated with enhanced metastasis in experimental models^{52,59}.

The pleiotropic effects of miRNA biology on cancer extend to virtually all acquired cancer traits, including cancer-associated changes in intracellular metabolism and the tissue microenvironment. For example, most cancer cells display alterations in glucose metabolism termed the Warburg effect⁶⁰. miRNAs may contribute to this metabolic switch because, in glioma cells, miR-451 controls cell proliferation, migration and responsiveness to glucose deprivation, thereby allowing the cells to survive metabolic stress⁶¹. The enhanced glutaminolysis observed in cancer cells can be partially explained by MYC-mediated repression

of miR-23a and miR-23b (ref. 62) (Fig. 2d). In some cases, the control of these cancer-related processes by miRNAs creates an opportunity for new therapeutic approaches. Hence, miR-132, which is present in the endothelium of tumours but not in normal human endothelium, induces neovascularization by inhibition of p120RasGAP, a negative regulator of KRAS⁶³. The delivery of a miR-132 inhibitor with nanoparticles that target the tumour vasculature suppresses angiogenesis in mice; this indicates there is a potential for the development of new antiangiogenic drugs. Further studies are likely to implicate miRNAs in the modulation of every tumour-associated pathway or trait.

Big lessons from mice

Much of what we have learnt concerning the functional contribution of miRNA biology to cancer development comes from studies in genetically engineered mice. These systems provide powerful tools for the genetic and biological study of miRNAs in an *in vivo* context, which is particularly important given the contextual activity of most miRNAs. In addition, owing to the ability of these models to recapitulate the behaviour of some human malignancies, they are useful in preclinical studies to evaluate new therapeutics.

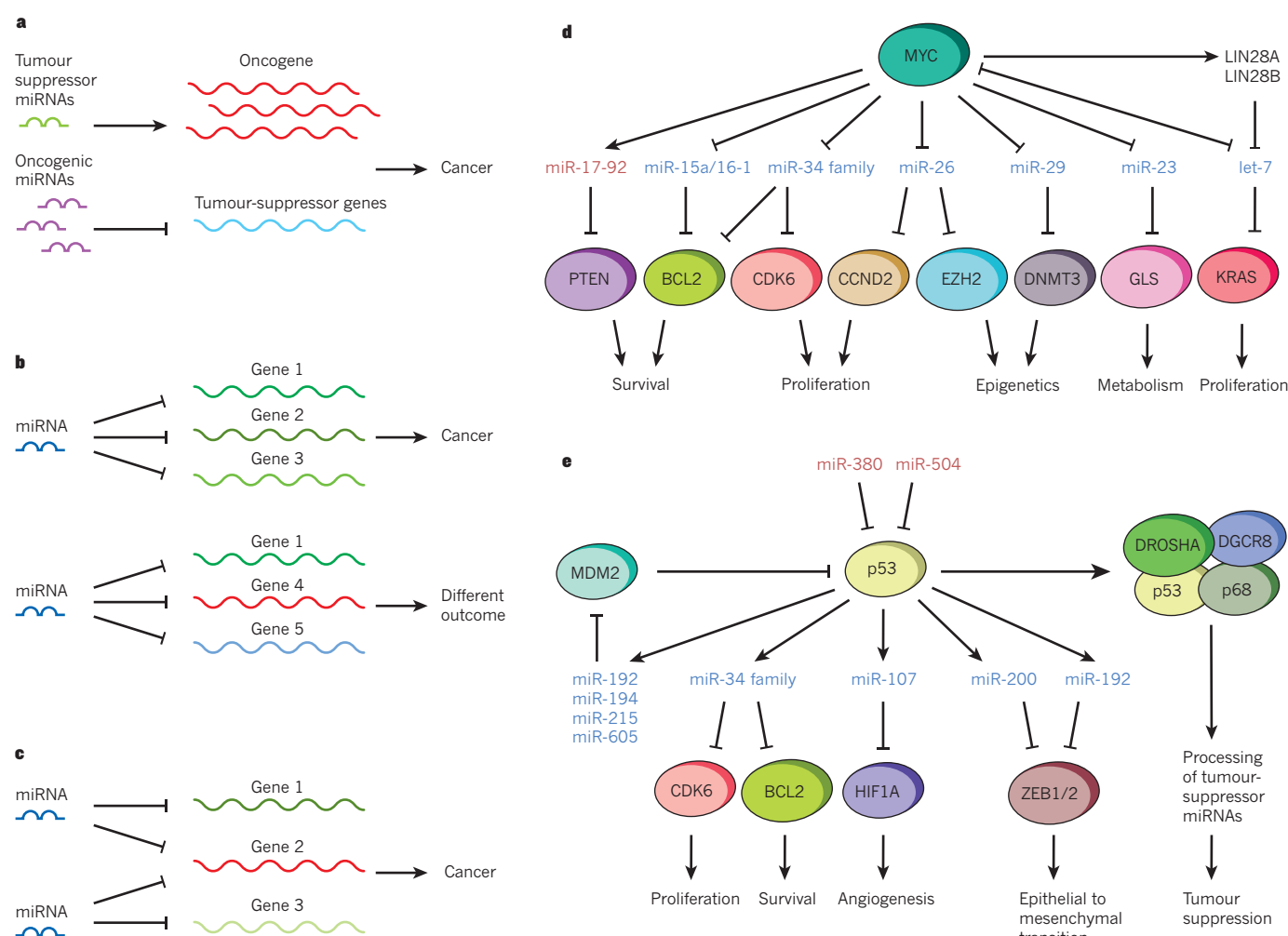


Figure 2 | Contribution of miRNAs to cancer pathways. **a**, Tumour-suppressor miRNAs, which repress oncogenes in healthy cells, are lost in cancer cells, leading to oncogene upregulation, whereas oncogenic miRNAs inhibit tumour-suppressor genes, giving rise to cancer. **b**, The presence of different target genes in different cell lines can modify the function of an miRNA, both in healthy cells and cancer cells, which can lead to the development of cancer or a different outcome. **c**, Two miRNAs can function together to regulate one or several pathways, which reinforces those pathways and can result in the development of cancer. **d**, The oncogene MYC can either

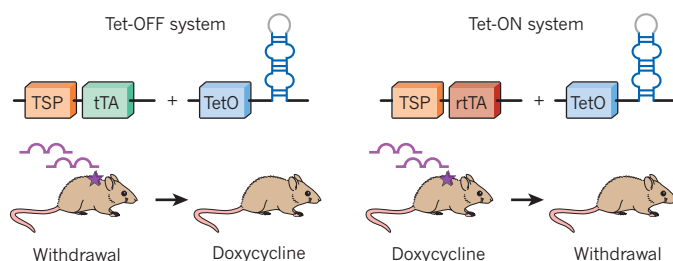
repress tumour-suppressor miRNAs (in blue) or activate oncogenic miRNAs (in red) and can therefore orchestrate several different pathways. MYC can repress let-7, directly, or indirectly, through LIN28 activation. Conversely, let-7 can also repress MYC, which closes the regulatory circle. **e**, Tumour suppressor p53 can regulate several tumour suppressor miRNAs (blue), activating different antitumoural pathways. The regulation of *MDM2* by some of these miRNAs leads to interesting feedforward loops. At the same time, p53 can be negatively regulated by oncogenic miRNAs (in red). In addition, p53 is involved in the biogenesis of several tumour suppressor miRNAs.

Perhaps the most widespread use of mice for characterizing miRNA biology in cancer is the validation of miRNAs that are altered in cancer cells, as bona fide oncogenes and tumour suppressor genes. As already mentioned, the first direct evidence that miRNAs have a function in cancer came from mouse models, in which it was shown that expression of the miR-17-92 cluster — which is amplified in some human B cell lymphomas — cooperates with Myc to promote B-cell lymphoma in mice⁸. Subsequent studies that have used genetically engineered or transplantation-based systems identified the relevant miRNA components, showing that the miR-19 family (including miR-19a and miR-19b) represents the most potent oncogenes in this cluster^{28,64,65}. Another example is miR-155 overexpression in the lymphoid compartment, which triggers B-cell leukaemia or a myeloproliferative disorder depending on the system used to drive expression of the transgene; this was the first example of a miRNA that initiates cancer in a transgenic setting^{66,67} (Table 1).

Gene targeting has been used extensively to delete miRNAs for the purpose of characterizing their physiological roles or action as candidate tumour suppressors. Gene targeting has suggested that miRNAs from similar families have redundant or compensatory functions, which has been shown for *C. elegans*⁶⁸. Ablation of the miR-15a and miR-16-1 cluster, which is often deleted in human chronic lymphocytic leukaemia, predisposes mice to B-cell lymphoproliferative disease⁶⁹ (Table 1). Importantly, the ability to produce mouse strains with different gene dosage through heterozygous or homozygous gene deletions has revealed that *Dicer1*, which if lost completely has a deleterious effect, can promote malignant phenotypes as a haploinsufficient tumour suppressor²³. Such a conclusion could not be formed from studies that examined only genomic data.

Conditional gene expression systems in mice have allowed researchers

a Inducible *in vivo* miRNA expression



b Inducible *in vivo* miRNA inhibition

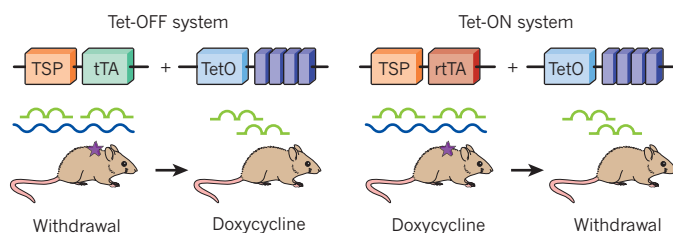


Figure 3 | *In vivo* miRNA expression or inhibition ‘à la carte’. a, Tetracycline (Tet)-mediated miRNA inactivation or activation by doxycycline administration using Tet-OFF, in which a tissue-specific promoter (TSP) is combined with a transactivator (tTA) to turn on expression of oncogenic miRNA (purple) and induce tumorigenesis (purple star) and subsequent tumour regression, revealing dependence on the oncogenic miRNA, or Tet-ON systems in which a reverse transactivator (rtTA) switches on oncogenic miRNA when the drug is applied. Drug withdrawal leads to tumour regression. b, Tet-mediated miRNA activation or inactivation by doxycycline administration using Tet-OFF or Tet-ON systems. miRNAs (green) can be inhibited by miRNA sponges (dark blue), with the same effects as miRNA expression, leading to tumorigenesis and subsequent tumour regression, which indicates a dependence on tumour-suppressor loss.

to determine cancer gene dependencies, as well as whether genes that initiate cancer also participate in tumour maintenance. In many cases, withdrawal of the initiating oncogenic transgene (or restoration of the deleted or lost tumour suppressor) leads to the collapse of the tumour; this validates the transgene or pathway that is controlled by these genes, as a therapeutic target. Similar studies have also been applied to miRNAs; for example, conditional expression of miR-21, which is broadly deregulated in cancer, can promote lymphomagenesis in mice⁷⁰ (Table 1). Silencing of miR-21 leads to disease regression, in part, by promoting apoptosis⁷⁰ (Fig. 3a). Likewise, the use of miRNA inhibitors (for example, antagomirs) directed against miR-21 can inhibit the proliferation of human cancer cells that overexpress miR-21 (ref. 71). Together, these studies suggest that miR-21 antagonists have the potential to be effective therapies for at least some cancers.

The development of new technology has meant that mouse models are increasingly used to study gene function on a large if not genome-wide scale, and miRNAs are at the forefront of this revolution. Recently, a vast collection of mouse embryonic stem-cell clones that harbour deletions that target 392 miRNA genes was generated⁷². This unique and valuable toolbox, termed ‘mirKO’, will allow the creation of mice that lack specific miRNAs, express mutant miRNAs or the study of their expression. In a converse strategy, a collection of embryonic stem cells engineered to inducibly express the vast majority of known miRNAs is in production (S.W.L., Y. Park and G. Hannon, manuscript in preparation) and will allow the *in vivo* validation of miRNAs as oncogenes or as anticancer therapies. With a different strategy, miRNA sponges (Fig. 3b), which are oligonucleotide constructs with multiple complementary miRNA binding sites in tandem, have already been used to deplete individual miRNAs in transgenic fruitflies, in transplanted breast cancer cells in mice and in a transgenic mouse model^{56,73,74}. Although these sponges provide a scalable strategy for miRNA loss-of-function studies, more work is needed to rule out off-target effects and assess their potency before conclusions can be made. However, the availability of such resources will help with the functional study of miRNAs in normal development and disease, and will be useful to the wider scientific community.

Finally, genetically engineered mouse models of human cancers are a testing ground for preclinical studies. For example, in Myc-induced liver tumours, miR-26 delivery by adeno-associated viruses suppresses tumorigenesis by inducing apoptosis⁷⁵. The increasing use of state-of-the-art mouse models is likely to uncover new *in vivo* functions, such as metastasis and angiogenesis, that otherwise would have remained hidden *in vitro*. They will also provide key preclinical systems for testing miRNA-based therapeutics.

Constructing and deconstructing cancer

The use of RNAi technology — a tool that exploits miRNA pathways — has revolutionized the study of gene function in mammalian systems and has provided a powerful means to investigate the function of any protein-coding gene. Experimental triggers of RNAi exploit different aspects of the pathway and result in the downregulation of gene expression through incorporation into the miRNA biogenesis machinery at different points⁷⁶. Small-interfering RNAs (siRNAs), which function at the level of Dicer1, can transiently and potentially lead to gene suppression; these RNAi triggers, or their variants, are probably the structural ‘scaffold’ for miRNA therapeutics (see the section miRNAs as drugs or drug targets).

Stable RNAi can be activated by the expression of miRNA mimetics, that are either the so-called stem loop short-hairpin RNAs (shRNAs) or shRNAs that incorporate a larger miRNA fold. One example of the latter is based on miR-30 (known as miR-30-based shRNAs or ‘shRNAmirs’). These shRNAs, as occurs naturally for many miRNAs, can be embedded in non-coding sequences of protein-coding transcripts or linked in tandem, which allows, for example, the linkage of the shRNA with a fluorescent reporter or the simultaneous knockdown of two different

genes^{77,78}. Advances in the shRNAir methodology have allowed the development of versatile vectors for the study of proliferation and survival genes, strategies for optimizing the potency of shRNAs, and rapid and effective systems for conditional shRNA expression in mice^{79–81}. The last of these, together with systems based on short stem-loop shRNAs⁸², could eventually allow the spatial, temporal and reversible control of any gene *in vivo*.

Regardless of the platform, RNAi technology provides an effective tool to investigate cancer phenotypes and identify therapeutic targets. For example, RNAi has been used to identify and characterize tumour-suppressor genes, which if inhibited promote cancer development. Early studies, using the same system that validated miR-17-92 as an oncogene, demonstrated that inhibition of *TP53* could produce phenotypes that were consistent with *TP53* loss⁸³. Later studies showed that tumour suppressors could be identified prospectively using *in vitro* and *in vivo* shRNA screens, (for examples see refs 84 and 85). By conditionally expressing shRNAs that target tumour suppressors in mice, tumour-suppressor function in advanced tumours can be re-established by silencing the shRNA⁸⁶. Tumour-suppressor reactivation leads to a marked (if not complete) tumour regression, which validates these pathways as therapeutic targets.

RNAi technology can be exploited more directly to identify genotype-specific cancer drug targets. Although there may be differences in the outcome of RNAi and small-molecule-mediated protein inhibition, siRNAs and shRNAs have been widely used to determine whether a candidate target is required for the proliferation of cancer cells. Moreover, the availability of RNAi libraries that target portions of, or all, the human genome allows genetic screens to identify 'synthetic lethal' genes, for which, if combined, the attenuation triggers the death of the cell. In principle, the identification of an RNAi target, the inhibition of which is selectively lethal to cells harbouring a particular oncogenic alteration, should identify cancer-specific targets. Such approaches have identified potential targets for *KRAS*-expressing tumours^{87–89} and leukaemias with deregulated *MYC* (ref. 90). Application of these approaches could potentially be complementary to the traditional drug-target discovery approach, and possibly a systematic way to identify the combination of therapies that will ultimately be needed to combat cancer.

miRNAs as drugs and drug targets

Despite advances in techniques to inhibit protein-coding genes using small molecules or biologicals, many cancers are unresponsive to the agents currently in use or become resistant to them; new and more creative approaches are therefore required for the treatment of cancer. Perhaps one of the most exciting opportunities that has arisen from our understanding of miRNA biology is the potential use of miRNA mimics or antagonists as therapeutics. Owing to the ability of miRNAs to simultaneously target multiple genes and pathways that are involved in cellular proliferation and survival³⁸, the targeting of a single miRNA can be a form of 'combination' therapy that could obstruct feedback and compensatory mechanisms that would otherwise limit the effectiveness of many therapies in current use. In addition, because miRNA expression is often altered in cancer cells, agents that modulate miRNA activity could potentially produce cancer-specific effects^{10,91,92}. Based on this, anticancer therapies that inhibit or enhance miRNA activity are being developed (Fig. 4). Evidence for this is shown by the inhibition of oncogenic miRNAs or the expression of tumour suppressor miRNAs in mice that harbour tumours, which have a significant effect on the outcome of cancer. Oncogenic miRNAs can be blocked by using antisense oligonucleotides, antagomirs, sponges or locked nucleic acid (LNA) constructs⁹³. The use of LNAs has achieved unexpected success *in vivo*, not only in mice but also for the treatment of hepatitis C in non-human primates⁹⁴. The downregulation of miR-122 can lead to a significant inhibition of replication of the hepatitis C virus. This inhibition is thought to decrease the risk of chronic hepatitis and hepatocellular carcinoma in patients who are hepatitis C-positive. Early clinical studies using SPC3649, an miR-122 antagonist, in

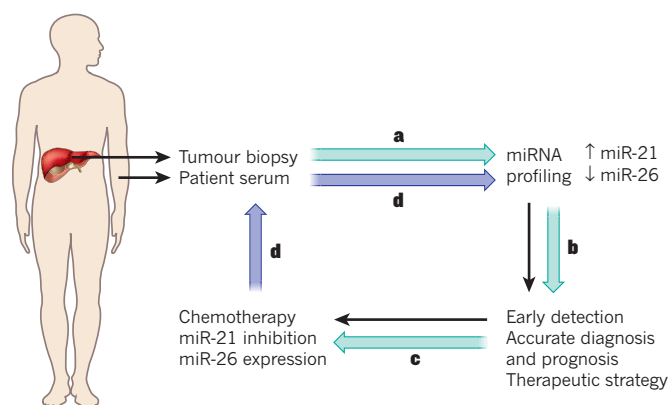


Figure 4 | Proposed scheme for the treatment of liver cancer with combined chemotherapy and miRNA-based therapy. **a**, miRNA expression profiles of potential patients could be assessed by measuring circulating miRNAs in patient serum or tumoral miRNAs from a biopsy. For example, miR-21 expression and miR-26 loss could be detected in serum and tumour samples. **b**, This profile could be used for early detection of cancer, accurate diagnosis and prognosis, and choosing the best therapeutic strategy. The best available chemotherapeutic option could be combined with miRNA-based therapy. **c**, The oncomiRs detected in miRNA profiling and those present in the tumour, such as miR-21, could be inhibited by using different strategies, such as locked nucleic acid constructs. By contrast, the expression of tumour-suppressor miRNAs downregulated in the tumour could be restored and miR-26 levels could be increased with miRNA mimics. **d**, After treatment, the patient could be checked for relapse by periodically studying circulating miRNAs from serum in a non-invasive manner. The presence of miR-21 could indicate a potential relapse, and treatment would resume (black arrows).

healthy individuals to assess toxicity will provide valuable information about pharmacokinetics and safety of the treatment. LNAs have been optimized to target miRNAs by reducing their molecular size and this, along with developing strategies for more efficient delivery, has increased their therapeutic potential⁹⁵. By contrast, another strategy involves the restoration of tumour-suppressor miRNA expression by synthetic miRNA mimics or viral delivery⁹³. Both of these approaches have yielded positive results in mouse models of cancer^{75,96}. Adeno-associated virus delivery of miRNAs or miRNA antagonists has the advantage of being efficient and, because the virus does not integrate into the genome, non-mutagenic. However, the delivery and safety of treatment needs to be improved before this approach can achieve widespread clinical use.

In principle, the use of miRNA mimetics as therapeutics would allow 'drugging the undruggable' or the therapeutic inhibition of virtually any human gene. If this were possible it would undoubtedly impact many diseases including cancer by allowing the targeting of oncogenic transcription factors that are difficult to inhibit through traditional medicinal chemistry⁹⁷. Furthermore, owing to the similar chemistry that is used to create drugs that target diverse molecules, the implementation of miRNA-based therapies could allow a more uniform drug development pipeline than is possible for more conventional treatments. Although experimental studies have validated the underlying biological impact of achieving miRNA modulation, there are still practical challenges that prevent the use of miRNA mimetics and antagonists clinically, including uncharacterized off-target effects, toxicities and poor agent-delivery. Concerning the last of these, most miRNA mimetics and antagonists rely on the delivery of molecules that mimic or inhibit the 'seed' sequence of an miRNA (typically molecules that consist of ≥6 nucleotides or related structures) across the plasma membrane — a particular challenge in the treatment of cancer, in which missing even a few cancer cells could lead to tumour relapse and progression. Extensive research is now focused on the viral and non-viral strategies required to meet this challenge, and results in the preclinical setting are promising^{75,94–96}. Despite the considerable hurdles that have to be overcome, it

seems likely that miRNAs will find a place alongside more conventional approaches for the treatment of cancer.

Perspectives

Since the discovery of miRNAs in model organisms, miRNAs have emerged as key regulators of normal development and a diversity of normal cellular processes. Given what we know now, it is not surprising that perturbations in miRNA biogenesis or expression can contribute to disease. In cancer, the effects of miRNA alteration can be widespread and profound, and they touch on virtually all aspects of the malignant phenotype. Yet, precisely how miRNAs regulate the expression of protein-coding genes is not completely understood, and the underlying mechanism remains an important basic-science question that will have a significant impact on our understanding of gene regulation and its alteration in disease. In addition, we still lack effective approaches to understand and predict miRNA targets. New strategies to identify and characterize the targets of individual miRNAs, and to determine how they function in combination to regulate specific targets, will be required to understand their action on cell physiology. Because miRNAs can also regulate other non-coding RNAs (for example, long non-coding RNAs), which have a role in cancer development and vice versa⁹⁸, these interactions will increase the complexity of gene regulation and are likely to produce regulatory processes that are currently hidden. Pioneering knowledge, gained through the study of miRNA function and regulation, will undoubtedly provide methodological and theoretical insights that will help in our understanding of the more recently identified non-coding RNA species.

Understanding miRNA biology and how it contributes to cancer development is not only an academic exercise, but also provides an opportunity for the generation of new ideas for diagnosis and treatment. RNAi-based technology has allowed sophisticated loss-of-function experiments that were previously impossible and has revealed therapeutic targets that, when inhibited, can lead to cancer cell elimination. In addition, miRNAs themselves are being used directly in the diagnosis of cancer and, in the future, will probably be exploited in therapy to identify drug targets or as the drug treatment. However, cost-effective miRNA profiling strategies and larger studies are needed to determine whether miRNA profiling provides an advantage for cancer classification compared with a more traditional approach. Although drugs that function as miRNA mimetics, antagonists or synthetic siRNAs form the core of what is fundamentally a new class of drugs that are capable of targeting molecules outside the range of traditional medicinal chemistry, their clinical implementation will require improvements in drug composition and delivery; these challenges lie outside the scope of molecular biology and instead involve the fields of chemistry and nanotechnology. Nevertheless, the successful development of these technologies could ultimately translate our understanding of miRNA biology in cancer into strategies for the control of cancer. ■

1. Bartel, D. P. MicroRNAs: target recognition and regulatory functions. *Cell* **136**, 215–233 (2009).
 2. Lee, R. C., Feinbaum, R. L. & Ambros, V. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* **75**, 843–854 (1993).
 3. Wightman, B., Ha, I. & Ruvkun, G. Post-transcriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern formation in *C. elegans*. *Cell* **75**, 855–862 (1993).
 4. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674 (2011).
 5. Calin, G. A. et al. Frequent deletions and down-regulation of micro-RNA genes *miR15* and *miR16* at 13q14 in chronic lymphocytic leukemia. *Proc. Natl Acad. Sci. USA* **99**, 15524–15529 (2002).
- This article reports miRNA deregulation in cancer and is the first evidence of the role of miRNAs in cancer.**
6. Lu, J. et al. MicroRNA expression profiles classify human cancers. *Nature* **435**, 834–838 (2005).
- This article systematically profiles miRNAs in cancer and demonstrates their potential as classifiers.**
7. O'Donnell, K. A., Wentzel, E. A., Zeller, K. I., Dang, C. V. & Mendell, J. T. c-Myc-regulated microRNAs modulate E2F1 expression. *Nature* **435**, 839–843 (2005).
 8. He, L. et al. A microRNA polycistron as a potential human oncogene. *Nature*

435, 828–833 (2005).

References 7 and 8 show, for the first time, that miRNAs can be actively involved in the MYC signalling pathway.

9. Calin, G. A. et al. Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. *Proc. Natl Acad. Sci. USA* **101**, 2999–3004 (2004).
 10. Saito, Y. et al. Specific activation of microRNA-127 with downregulation of the proto-oncogene *BCL6* by chromatin-modifying drugs in human cancer cells. *Cancer Cell* **9**, 435–443 (2006).
 11. Mayr, C., Hemann, M. T. & Bartel, D. P. Disrupting the pairing between *let-7* and *Hmga2* enhances oncogenic transformation. *Science* **315**, 1576–1579 (2007).
 12. Veronese, A. et al. Mutated β -catenin evades a microRNA-dependent regulatory loop. *Proc. Natl Acad. Sci. USA* **108**, 4840–4845 (2011).
 13. Diederichs, S. & Haber, D. A. Sequence variations of microRNAs in human cancer: alterations in predicted secondary structure do not affect processing. *Cancer Res.* **66**, 6097–6104 (2006).
 14. Kuchenbauer, F. et al. In-depth characterization of the microRNA transcriptome in a leukemia progression model. *Genome Res.* **18**, 1787–1797 (2008).
 15. Yanaihara, N. et al. Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. *Cancer Cell* **9**, 189–198 (2006).
 16. Calin, G. A. et al. A microRNA signature associated with prognosis and progression in chronic lymphocytic leukemia. *N. Engl. J. Med.* **353**, 1793–1801 (2005).
 17. Rosenfeld, N. et al. MicroRNAs accurately identify cancer tissue origin. *Nature Biotechnol.* **26**, 462–469 (2008).
 18. Xi, Y. et al. Systematic analysis of microRNA expression of RNA extracted from fresh frozen and formalin-fixed paraffin-embedded samples. *RNA* **13**, 1668–1674 (2007).
 19. Mitchell, P. S. et al. Circulating microRNAs as stable blood-based markers for cancer detection. *Proc. Natl Acad. Sci. USA* **105**, 10513–10518 (2008).
 20. Chang, T. C. et al. Widespread microRNA repression by Myc contributes to tumorigenesis. *Nature Genet.* **40**, 43–50 (2008).
 21. Thomson, J. M. et al. Extensive post-transcriptional regulation of microRNAs and its implications for cancer. *Genes Dev.* **20**, 2202–2207 (2006).
 22. Kumar, M. S., Lu, J., Mercer, K. L., Golub, T. R. & Jacks, T. Impaired microRNA processing enhances cellular transformation and tumorigenesis. *Nature Genet.* **39**, 673–677 (2007).
 23. Kumar, M. S. et al. *Dicer1* functions as a haploinsufficient tumor suppressor. *Genes Dev.* **23**, 2700–2704 (2009).
 24. Merritt, W. M. et al. *Dicer*, *Drosha*, and outcomes in patients with ovarian cancer. *N. Engl. J. Med.* **359**, 2641–2650 (2008).
 25. Melo, S. A. et al. A *TARBP2* mutation in human cancer impairs microRNA processing and *DICER1* function. *Nature Genet.* **41**, 365–370 (2009).
 26. Melo, S. A. et al. A genetic defect in exportin-5 traps precursor microRNAs in the nucleus of cancer cells. *Cancer Cell* **18**, 303–315 (2010).
 27. Newman, M. A. & Hammond, S. M. Emerging paradigms of regulated microRNA processing. *Genes Dev.* **24**, 1086–1092 (2010).
 28. Mavrakis, K. J. et al. Genome-wide RNA-mediated interference screen identifies miR-19 targets in Notch-induced T-cell acute lymphoblastic leukaemia. *Nature Cell Biol.* **12**, 372–379 (2010).
 29. Portela, A. & Esteller, M. Epigenetic modifications and human disease. *Nature Biotechnol.* **28**, 1057–1068 (2010).
 30. Cao, Q. et al. Coordinated regulation of Polycomb Group complexes through microRNAs in cancer. *Cancer Cell* **20**, 187–199 (2011).
 31. Fabbri, M. et al. MicroRNA-29 family reverts aberrant methylation in lung cancer by targeting DNA methyltransferases 3A and 3B. *Proc. Natl Acad. Sci. USA* **104**, 15805–15810 (2007).
 32. Varambally, S. et al. Genomic loss of microRNA-101 leads to overexpression of histone methyltransferase EZH2 in cancer. *Science* **322**, 1695–1699 (2008).
 33. Hwang, H. W., Wentzel, E. A. & Mendell, J. T. A hexanucleotide element directs microRNA nuclear import. *Science* **315**, 97–100 (2007).
 34. Khraiweh, B. et al. Transcriptional control of gene expression by microRNAs. *Cell* **140**, 111–122 (2010).
 35. Gebeshuber, C. A., Zatloukal, K. & Martinez, J. miR-29a suppresses tristetraprolin, which is a regulator of epithelial polarity and metastasis. *EMBO Rep.* **10**, 400–405 (2009).
 36. Small, E. M. & Olson, E. N. Pervasive roles of microRNAs in cardiovascular biology. *Nature* **469**, 336–342 (2011).
 37. Bueno, M. J. et al. Combinatorial effects of microRNAs to suppress the Myc oncogenic pathway. *Blood* **117**, 6255–6266 (2011).
 38. Bui, T. V. & Mendell, J. T. Myc: maestro of microRNAs. *Genes Cancer* **1**, 568–575 (2010).
 39. Dews, M. et al. Augmentation of tumor angiogenesis by a Myc-activated microRNA cluster. *Nature Genet.* **38**, 1060–1065 (2006).
 40. Cairo, S. et al. Stem cell-like micro-RNA signature driven by Myc in aggressive liver cancer. *Proc. Natl Acad. Sci. USA* **107**, 20471–20476 (2010).
 41. Kent, O. A. et al. Repression of the miR-143/145 cluster by oncogenic Ras initiates a tumor-promoting feed-forward pathway. *Genes Dev.* **24**, 2754–2759 (2010).
 42. Johnson, S. M. et al. RAS is regulated by the *let-7* microRNA family. *Cell* **120**, 635–647 (2005).
- This article reports the first evidence of an oncogene, KRAS, being targeted by an miRNA.**
43. He, L., He, X., Lowe, S. W. & Hannon, G. J. microRNAs join the p53 network—another piece in the tumour-suppression puzzle. *Nature Rev. Cancer* **7**, 819–822 (2007).

This comprehensive review describes the regulation of the miR-34 family by the tumour suppressor p53.

44. Pichiotti, F. *et al.* Downregulation of p53-inducible microRNAs 192, 194, and 215 impairs the p53/MDM2 autoregulatory loop in multiple myeloma development. *Cancer Cell* **18**, 367–381 (2010).
45. Xiao, J., Lin, H., Luo, X. & Wang, Z. miR-605 joins p53 network to form a p53:miR-605:MDM2 positive feedback loop in response to stress. *EMBO J.* **30**, 524–532 (2011).
46. Yamakuchi, M. *et al.* P53-induced microRNA-107 inhibits HIF-1 and tumor angiogenesis. *Proc. Natl Acad. Sci. USA* **107**, 6334–6339 (2010).
47. Chang, C. J. *et al.* p53 regulates epithelial-mesenchymal transition and stem cell properties through modulating miRNAs. *Nature Cell Biol.* **13**, 317–323 (2011).
48. Kim, T. *et al.* p53 regulates epithelial-mesenchymal transition through microRNAs targeting ZEB1 and ZEB2. *J. Exp. Med.* **208**, 875–883 (2011).
49. Swarbrick, M. *et al.* miR-380-5p represses p53 to control cellular survival and is associated with poor outcome in MYCN-amplified neuroblastoma. *Nature Med.* **16**, 1134–1140 (2010).
50. Hu, W. *et al.* Negative regulation of tumor suppressor p53 by microRNA miR-504. *Mol. Cell* **38**, 689–699 (2010).
51. Suzuki, H. I. *et al.* Modulation of microRNA processing by p53. *Nature* **460**, 529–533 (2009).
52. Su, X. *et al.* TAp63 suppresses metastasis through coordinate regulation of Dicer and miRNAs. *Nature* **467**, 986–990 (2010).
53. Ma, L., Teruya-Feldstein, J. & Weinberg, R. A. Tumour invasion and metastasis initiated by microRNA-10b in breast cancer. *Nature* **449**, 682–688 (2007).
- This study demonstrates for the first time that miRNAs are involved in tumour invasion and metastasis.**
54. Tavazoie, S. F. *et al.* Endogenous human microRNAs that suppress breast cancer metastasis. *Nature* **451**, 147–152 (2008).
55. Ma, L. *et al.* miR-9, a MYC/MYCIN-activated microRNA, regulates E-cadherin and cancer metastasis. *Nature Cell Biol.* **12**, 247–256 (2010).
56. Valastyan, S. *et al.* A pleiotropically acting microRNA, miR-31, inhibits breast cancer metastasis. *Cell* **137**, 1032–1046 (2009).
57. Cano, A. & Nieto, M. A. Non-coding RNAs take centre stage in epithelial-to-mesenchymal transition. *Trends Cell Biol.* **18**, 357–359 (2008).
58. Korpai, M. *et al.* Direct targeting of Sec23a by miR-200s influences cancer cell secretome and promotes metastatic colonization. *Nature Med.* **17**, 1101–1108 (2011).
59. Martello, G. *et al.* A microRNA targeting Dicer for metastasis control. *Cell* **141**, 1195–1207 (2010).
60. Vander Heiden, M. G., Cantley, L. C. & Thompson, C. B. Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science* **324**, 1029–1033 (2009).
61. Godlewski, J. *et al.* MicroRNA-451 regulates LKB1/AMPK signaling and allows adaptation to metabolic stress in glioma cells. *Mol. Cell* **37**, 620–632 (2010).
62. Gao, P. *et al.* c-Myc suppression of miR-23a/b enhances mitochondrial glutaminase expression and glutamine metabolism. *Nature* **458**, 762–765 (2009).
63. Anand, S. *et al.* MicroRNA-132-mediated loss of p120RasGAP activates the endothelium to facilitate pathological angiogenesis. *Nature Med.* **16**, 909–914 (2010).
64. Mu, P. *et al.* Genetic dissection of the miR-17–92 cluster of microRNAs in Myc-induced B-cell lymphomas. *Genes Dev.* **23**, 2806–2811 (2009).
65. Olive, V. *et al.* miR-19 is a key oncogenic component of miR-17–92. *Genes Dev.* **23**, 2839–2849 (2009).
66. Costinean, S. *et al.* Pre-B cell proliferation and lymphoblastic leukemia/high-grade lymphoma in E(mu)-miR155 transgenic mice. *Proc. Natl Acad. Sci. USA* **103**, 7024–7029 (2006).
- This article reports overexpression of a single miRNA can cause cancer in vivo.**
67. O'Connell, R. M. *et al.* Sustained expression of microRNA-155 in hematopoietic stem cells causes a myeloproliferative disorder. *J. Exp. Med.* **205**, 585–594 (2008).
68. Miska, E. A. *et al.* Most *Caenorhabditis elegans* microRNAs are individually not essential for development or viability. *PLoS Genet.* **3**, e215 (2007).
69. Klein, U. *et al.* The DLEU2/miR-15a/16-1 cluster controls B cell proliferation and its deletion leads to chronic lymphocytic leukemia. *Cancer Cell* **17**, 28–40 (2010).
70. Medina, P. P., Nolde, M. & Slack, F. J. OncomiR addition in an *in vivo* model of microRNA-21-induced pre-B-cell lymphoma. *Nature* **467**, 86–90 (2010).
71. Chan, J. A., Krichevsky, A. M. & Kosik, K. S. MicroRNA-21 is an antiapoptotic factor in human glioblastoma cells. *Cancer Res.* **65**, 6029–6033 (2005).
72. Prosser, H. M., Koike-Yusa, H., Cooper, J. D., Law, F. C. & Bradley, A. A resource of vectors and ES cells for targeted deletion of microRNAs in mice. *Nature Biotechnol.* **29**, 840–845 (2011).
73. Loya, C. M., Lu, C. S., Van Vactor, D. & Fulga, T. A. Transgenic microRNA inhibition with spatiotemporal specificity in intact organisms. *Nature Methods* **6**, 897–903 (2009).
74. Zhu, Q. *et al.* A sponge transgenic mouse model reveals important roles for the miRNA-183/96/182 cluster in post-mitotic photoreceptors of the retina.

J. Biol. Chem. **2865**, 31749–31760 (2011).

This article reports the development of the first sponge transgenic mouse that allows *in vivo* inhibition of one or several miRNAs.

75. Kota, J. *et al.* Therapeutic microRNA delivery suppresses tumorigenesis in a murine liver cancer model. *Cell* **137**, 1005–1017 (2009).
- This article uses adenovirus-associated vectors to deliver miRNAs to the liver and treat cancer.**
76. Czech, B. & Hannon, G. J. Small RNA sorting: matchmaking for Argonautes. *Nature Rev. Genet.* **12**, 19–31 (2011).
77. Chicas, A. *et al.* Dissecting the unique role of the retinoblastoma tumor suppressor during cellular senescence. *Cancer Cell* **17**, 376–387 (2010).
78. Stegmeier, F., Hu, G., Rickles, R. J., Hannon, G. J. & Elledge, S. J. A lentiviral microRNA-based system for single-copy polymerase II-regulated RNA interference in mammalian cells. *Proc. Natl Acad. Sci. USA* **102**, 13212–13217 (2005).
79. Zuber, J. *et al.* Toolkit for evaluating genes required for proliferation and survival using tetracycline-regulated RNAi. *Nature Biotechnol.* **29**, 79–83 (2010).
80. Fellmann, C. *et al.* Functional identification of optimized RNAi triggers using a massively parallel sensor assay. *Mol. Cell* **41**, 733–746 (2011).
81. Premsrirut, P. K. *et al.* A rapid and scalable system for studying gene function in mice using conditional RNA interference. *Cell* **145**, 145–158 (2011).
82. Seibler, J. *et al.* Reversible gene knockdown in mice using a tight, inducible shRNA expression system. *Nucleic Acids Res.* **35**, e54 (2007).
83. Hemann, M. T. *et al.* An epi-allelic series of p53 hypomorphs created by stable RNAi produces distinct tumor phenotypes *in vivo*. *Nature Genet.* **33**, 396–400 (2003).
84. Zender, L. *et al.* An oncogenomics-based *in vivo* RNAi screen identifies tumor suppressors in liver cancer. *Cell* **135**, 852–864 (2008).
85. Westbrook, T. F. *et al.* A genetic screen for candidate tumor suppressors identifies REST. *Cell* **121**, 837–848 (2005).
86. Xue, W. *et al.* Senescence and tumour clearance is triggered by p53 restoration in murine liver carcinomas. *Nature* **445**, 656–660 (2007).
87. Luo, J. *et al.* A genome-wide RNAi screen identifies multiple synthetic lethal interactions with the Ras oncogene. *Cell* **137**, 835–848 (2009).
88. Scholl, C. *et al.* Synthetic lethal interaction between oncogenic KRAS dependency and STK33 suppression in human cancer cells. *Cell* **137**, 821–834 (2009).
89. Barbie, D. A. *et al.* Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* **462**, 108–112 (2009).
90. Zuber, J. *et al.* RNAi screen identifies Brd4 as a therapeutic target in acute myeloid leukaemia. *Nature* **478**, 524–528 (2011).
91. Gumireddy, K. *et al.* Small-molecule inhibitors of microRNA miR-21 function. *Angew. Chem. Int. Ed. Engl.* **47**, 7482–7484 (2008).
92. Melo, S. *et al.* Small molecule enoxacin is a cancer-specific growth inhibitor that acts by enhancing TAR RNA-binding protein 2-mediated microRNA processing. *Proc. Natl Acad. Sci. USA* **108**, 4394–4399 (2011).
93. Garzon, R., Marcucci, G. & Croce, C. M. Targeting microRNAs in cancer: rationale, strategies and challenges. *Nature Rev. Drug Discov.* **9**, 775–789 (2010).
94. Lanford, R. E. *et al.* Therapeutic silencing of microRNA-122 in primates with chronic hepatitis C virus infection. *Science* **327**, 198–201 (2009).
95. Obad, S. *et al.* Silencing of microRNA families by seed-targeting tiny LNAs. *Nature Genet.* **43**, 371–378 (2011).
96. Bonci, D. *et al.* The miR-15a-miR-16-1 cluster controls prostate cancer by targeting multiple oncogenic activities. *Nature Med.* **14**, 1271–1277 (2008).
97. Kumar, M. S. *et al.* Suppression of non-small cell lung tumor development by the let-7 microRNA family. *Proc. Natl Acad. Sci. USA* **105**, 3903–3908 (2008).
98. Poliseno, L. *et al.* A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature* **465**, 1033–1038 (2010).
- This elegant study shows how mRNA from genes and pseudogenes can compete for the binding of miRNAs, unveiling the complexity of miRNA regulatory networks.**
99. Ruby, J. G., Jan, C. H. & Bartel, D. P. Intronic microRNA precursors that bypass Drosha processing. *Nature* **448**, 83–86 (2007).
100. Cheloufi, S., Dos Santos, C. O., Chong, M. M. & Hannon, G. J. A dicer-independent miRNA biogenesis pathway that requires Ago catalysis. *Nature* **465**, 584–589 (2010).

Acknowledgements We apologize to all colleagues whose work could not be cited owing to space restrictions. We thank L. Dow, A. Ventura, A. Saborowski and V. Aranda for their comments on the manuscript, and G. Hannon and L. He for the many discussions. A.L. is supported by an EMBO Long-Term Fellowship. S.W.L. is a Howard Hughes Medical Institute investigator.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence should be addressed to S.W.L. (lowes@mskcc.org).

Eutrophication causes speciation reversal in whitefish adaptive radiations

P. Vonlanthen^{1,2}, D. Bittner^{2,3}, A. G. Hudson^{1,2}, K. A. Young^{2,4}, R. Müller², B. Lundsgaard-Hansen^{1,2}, D. Roy^{2,5}, S. Di Piazza^{1,2}, C. R. Lurgiader⁶ & O. Seehausen^{1,2}

Species diversity can be lost through two different but potentially interacting extinction processes: demographic decline and speciation reversal through introgressive hybridization. To investigate the relative contribution of these processes, we analysed historical and contemporary data of replicate whitefish radiations from 17 pre-alpine European lakes and reconstructed changes in genetic species differentiation through time using historical samples. Here we provide evidence that species diversity evolved in response to ecological opportunity, and that eutrophication, by diminishing this opportunity, has driven extinctions through speciation reversal and demographic decline. Across the radiations, the magnitude of eutrophication explains the pattern of species loss and levels of genetic and functional distinctiveness among remaining species. We argue that extinction by speciation reversal may be more widespread than currently appreciated. Preventing such extinctions will require that conservation efforts not only target existing species but identify and protect the ecological and evolutionary processes that generate and maintain species.

Effectively counteracting the biodiversity crisis requires identifying and protecting the ecological and evolutionary processes that generate and maintain diversity^{1,2}. Species can go extinct through two distinct but potentially interacting processes. In the first, demographic decline results in population extirpation and eventually the total extinction of the species. In the second, introgressive hybridization erodes differentiation until species collapse into a hybrid swarm³. A special case of introgressive hybridization is speciation reversal⁴, in which changes in selection regimes increase gene flow between sympatric species, thus eroding genetic and ecological differences. Speciation reversal may be particularly important in adaptive radiations with recently diverged sympatric species that lack strong intrinsic post-zygotic isolation^{5–8}.

Adaptive radiation is the evolution of ecological diversity in rapidly speciating lineages⁹. It is often characterized by ‘ecological speciation’, in which traits that are under divergent natural selection, or those genetically correlated with them, contribute to reproductive isolation^{9–13}. When reproductive isolation between ecologically differentiated populations is maintained by the temporal and spatial clustering of breeding aggregations, adaptive radiation occurs through the correlated partitioning of ecological and reproductive niche spaces. Because intrinsic post-zygotic isolation is typically weak during adaptive radiation¹², environmental changes that reduce niche space and relax the selective forces maintaining reproductive isolation^{14,15} can lead to extinction by speciation reversal^{4–8}.

Fish of post-glacial lakes are model systems for studying adaptive radiation owing to their recent origins and repeated patterns of diversification in independent lineages^{16–18}. These radiations are characterized by the correlated partitioning of ecological and reproductive niche spaces^{16,19,20}. In the European Alps, at least 25 lakes harbour 1 to 5 whitefish species (*Coregonus* spp.)^{18,21} (Fig. 1a and Supplementary Table 1). For 17 of these lakes, 13 of which contain multiple sympatric species, the whitefish diversity was described by Steinmann 60 years

ago²². This diversity has arisen since deglaciation within nine hydrologically independent lake systems¹⁸.

Reproductive isolation in central European whitefish radiations is maintained mainly by pre-zygotic mechanisms (divergence in spawning depth²³, time, possibly mate choice (B. Lundsgaard-Hansen et al. unpublished data) and extrinsic rather than intrinsic post-zygotic mechanisms²⁴. Generally, large-bodied species with few, widely spaced gill-rakers (benthic invertebrate feeders), spawn in winter in shallow littoral habitats, whereas small-bodied species with many densely spaced rakers (zooplankton feeders), spawn in deeper water in winter or summer. Exceptions to this rule are profundal summer-spawning species with very low numbers of gill-rakers that exist in Lake Thun and existed in Lake Constance²². Summer-spawning species choose cold and well-oxygenated spawning habitats below the thermocline (>20 m in depth). Eggs settle onto the lake-floor sediment and require an oxygenated water–sediment interface to develop and hatch²⁵. Because whitefish use most of the lacustrine habitats, and because of their large biomass and ecological diversity, they are keystone species in the ecosystems of pre-alpine lakes, which are commonly referred to as whitefish lakes.

Although eutrophication threatens lake ecosystems worldwide^{26,27}, the manner and mechanisms by which it has affected adaptive radiations, and whitefish in particular, remain unclear^{22,23}. Many Swiss whitefish lakes lie in densely populated areas and were subjected to high nutrient inputs in the twentieth century, a fact that led Steinmann to suggest in 1950 that eutrophication was the cause of the extinction of eight whitefish populations²². By the 1970s, eutrophication had increased primary production in all Swiss lakes (Fig. 2d and Supplementary Fig. 1). The associated increase in microbial decomposition rates resulted in oxygen depletion at the water–sediment interface, especially below the thermocline, leading to reduction or complete failure of whitefish recruitment²⁵. Eutrophication also affected the biomass and diversity of zooplankton (Supplementary Fig. 2) and

¹Division of Aquatic Ecology and Evolution, Institute of Ecology and Evolution, University of Bern, Baltzerstrasse 6, CH-3012 Bern, Switzerland. ²Department of Fish Ecology and Evolution, Centre of Ecology, Evolution and Biogeochemistry, EAWAG Swiss Federal Institute of Aquatic Science and Technology, Seestrasse 79, CH-6047 Kastanienbaum, Switzerland. ³Computational and Molecular Population Genetics (CMPG) Laboratory, Institute of Ecology and Evolution, University of Bern, Baltzerstrasse 6, CH-3012 Bern, Switzerland. ⁴Environment Agency, Cambria House, Newport Road, Cardiff CF24 0TP, UK. ⁵Department of Biology, Dalhousie University, 1355 Oxford Street, Halifax, Nova Scotia B3H 4R2, Canada. ⁶Institute of Clinical Chemistry, Inselspital University Hospital and University of Bern, Inselspital, CH-3010 Bern, Switzerland.

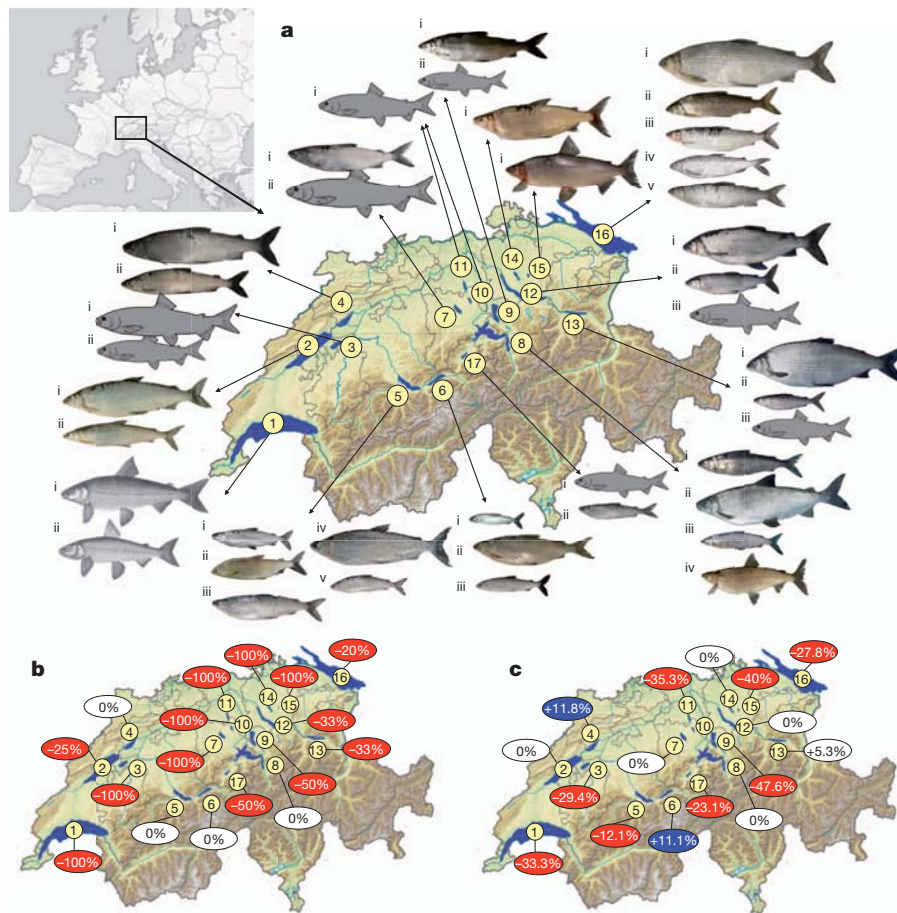


Figure 1 | Distribution of historical whitefish diversity and recent diversity loss. **a**, Whitefish species diversity in Swiss lakes (numbered as in Table 1, fish are named in Supplementary Information; for details of taxonomy see quantification of whitefish diversities in Supplementary Information). **b**, Species richness change in 17 lakes. **c**, Functional diversity change in 16 lakes.

probably of benthic invertebrates^{28–30}, thus altering the ecological and reproductive niche spaces that were associated with whitefish radiations. Improved sewage treatment and phosphorus management have allowed some lakes to return to near their natural trophic state (Fig. 2d). However, in other lakes, the sediment–water interface remains anoxic and zooplankton biomass is higher than before eutrophication²⁸.

We suspected that loss of deep spawning habitat weakened reproductive isolation, and that at the same time, increased productivity led to an increase of zooplankton density at the expense of zooplankton diversity (Supplementary Fig. 2), whereas the associated hypoxia probably led to loss of zoobenthos density in the profundal zone³¹. By disproportionately affecting the availability of one type of prey more than the other along the principal axis of whitefish feeding divergence, eutrophication probably changed the shape of the adaptive landscape from multimodal towards unimodal or flat, thus relaxing divergent selection. We therefore proposed that eutrophication caused speciation reversal in addition to demographic decline. We show that the speciation reversal hypothesis is supported by historical and contemporary patterns of diversity across lakes and by changes through time in genetic and phenotypic distinctiveness of sympatric species.

Diversity loss in polluted lakes

Most whitefish assemblages have lower species and functional diversity today than historically (Fig. 1, Table 1, Supplementary Table 1 and quantification of whitefish diversity in Supplementary Information). On average, species richness has decreased by 38% (Wilcoxon test $N = 17$, $V = 91$, $P < 0.001$), functional diversity (range in gill-raker

Red ellipses, more than 10% diversity loss; white ellipses, little or no change; blue ellipses, increase in diversity of more than 10%. The observed functional diversity increase in Lake Brienz is due to the presence of one species (*C. sp. 'Balchen'*) that Steinmann was unaware of²².

numbers) by 14% ($N = 16$, $V = 60$, $P = 0.018$) and the difference between sympatric species in gill-raker mean counts by 28% (Welch's t -test $N = 8$, $t = 7.79$, d.f. = 7, $P < 0.001$). Declines in species richness were explained by eutrophication level (linear regression $N = 17$, $R^2 = 0.50$, $P < 0.001$; Fig. 2a and Supplementary Table 2). Reductions in gill-raker count range were poorly predicted by eutrophication, probably because some variation is retained in hybrid swarms and stocking programmes have maintained some diversity even in the most polluted lakes²⁵ (Table 1 and Supplementary Table 1). Eutrophication reduced the oxygenated depth (depth range with $O_2 > 2.5 \text{ mg l}^{-1}$; see Supplementary Information) across 16 lakes (Supplementary Fig. 3). Egg survival was measured in a subset of those lakes and was found to decrease with nutrient load ($N = 12$, $R^2 = 0.45$, $P = 0.010$; Fig. 3f) and was close to zero once the maximum phosphorus exceeded $150 \mu\text{g l}^{-1}$.

Predicting the origin and loss of diversity

Because available depth affects the diversity of spawning and feeding habitats^{22,23}, and because all lakes were oxygenated to the greatest depths before eutrophication, we expected maximum lake depth (D_{max}) to predict pre-eutrophication diversity. By contrast, we expected maximum phosphorus concentration (P_{max}) and minimum oxygenated depth ($D_{O,\text{min}}$) during eutrophication to predict patterns of contemporary diversity (Table 1).

Maximum lake depth does indeed predict historical species richness ($N = 17$, $R^2 = 0.48$, $P = 0.001$, Fig. 3a) and the use of vulnerable reproductive niches. This pattern held when tested with evolutionarily independent lineages from hydrologically isolated lake

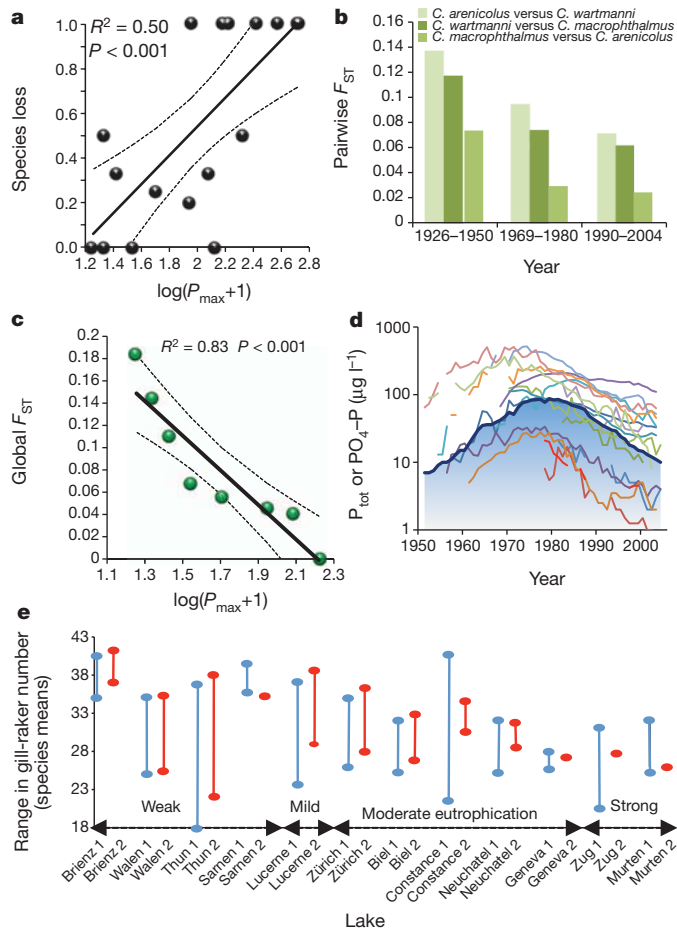


Figure 2 | Diversity loss through speciation reversal. **a**, Species diversity loss regressed against the maximum phosphorus concentration, P_{\max} ($\mu\text{g l}^{-1}$). **b**, The pairwise F_{ST} values among three *Coregonus* species from Lake Constance observed through time. **c**, The global genetic differentiation among species within each lake plotted against the maximum phosphorus concentration. Lake details are given in Supplementary Fig. 1. **d**, Fifty-year trends in phosphorus concentration from our study lakes are included. Lake Constance is highlighted as a blue gradient surface. Lake details are given in Supplementary Fig. 1. **e**, Ranges of species means in gill-raker counts for each lake, prior to (historical; 1, shown in blue) and after (contemporary; 2, shown in red) pollution. Lakes are arranged from weakly to strongly polluted. For panels **a** and **c** the dashed lines represent the 95% confidence intervals for the regression line.

groups^{18,22} as the unit of observation ($N = 9$, $R^2 = 0.51$, $P = 0.019$). Historical functional diversity also increased with maximum lake depth ($N = 17$, $R^2 = 0.30$, $P = 0.013$; Fig. 3b). Lakes that historically harboured summer- and deep-spawning species are significantly deeper than lakes that did not ($N_{\text{Summer}} = 8$, $N_{\text{NoSummer}} = 9$; $t = 3.25$, d.f. = 12.97, $P = 0.006$; $N_{\text{Deep}} = 11$, $N_{\text{NoDeep}} = 6$; $t = 5.05$, d.f. = 14.36, $P < 0.001$).

Oxygenated depth during eutrophication predicts contemporary species richness and functional diversity slightly better than does maximum lake depth, although the difference is not significant (difference in Akaike's corrected information criterion (ΔAICc) = 1.96 and 1.91; see regression model selection in Supplementary Information) (species richness: $N = 17$, $R^2 = 0.55$, $P < 0.001$, Fig. 3c, versus $R^2 = 0.49$, $P < 0.001$; functional diversity: $N = 16$, $R^2 = 0.40$, $P = 0.005$, Fig. 3d, versus $R^2 = 0.32$, $P = 0.013$). This was also true for historical species richness and functional diversity, but oxygenated depth explained slightly more of the variance in contemporary than in historical diversity (Supplementary Table 2). Moreover, lakes that lost summer- or deep-spawning species were more eutrophied than those that retained these species ($N_{\text{SummerLoss}} = 3$, $N_{\text{SummerNoLoss}} = 5$; $t = 3.04$, d.f. = 5.99, $P = 0.023$; $N_{\text{DeepLoss}} = 3$,

$N_{\text{DeepNoLoss}} = 8$; $t = 2.98$, d.f. = 7.13, $P = 0.020$), whereas maximum depth was not different between these lakes ($N_{\text{SummerLoss}} = 3$, $N_{\text{SummerNoLoss}} = 5$; $t = 0.44$, d.f. = 5.90, $P = 0.675$; $N_{\text{DeepLoss}} = 3$, $N_{\text{DeepNoLoss}} = 8$; $t = -0.301$, d.f. = 2.27, $P = 0.783$).

Among lakes, the contemporary number of genetically differentiated species (see Methods) is best predicted by maximum depth ($N = 8$, $R^2 = 0.50$, $P = 0.031$; Fig. 3e). The level of genetic differentiation among species, on the other hand, is predicted by the severity of eutrophication, to which it is strongly negatively correlated ($N = 8$, $R^2 = 0.83$, $P < 0.001$; Fig. 2c). In combination with the previous results, these data suggest that the depth-mediated legacy of adaptive radiation has been modified by speciation reversal driven by eutrophication.

Species loss through speciation reversal

If extinction resulted from demographic decline, pairwise genetic differentiation among contemporary species at neutral markers (measured using the fixation index (F_{ST})) would remain unchanged or increase owing to genetic drift as effective population sizes declined³². By contrast, extinction by speciation reversal should involve declines in pairwise F_{ST} values among extant species⁶. Lake Constance suffered eutrophication, but phosphorus concentrations never exceeded $150 \mu\text{g l}^{-1}$ at which egg development fails even in shallow waters (Table 1). We extracted DNA from samples of all species collected before (1926–50, $P_{\max} < 10 \mu\text{g l}^{-1}$), during (1969–80, $P_{\max} = 87 \mu\text{g l}^{-1}$) and after (1990–2004, $P_{\max} = 39 \mu\text{g l}^{-1}$) peak eutrophication. Genetic cluster analysis identified four species, with all four being well represented in pre-eutrophication scale samples. Out of all of the post-eutrophication samples, only five individuals were assigned to the now extinct summer- and deep-spawning *Coregonus gutturosus* (Supplementary Table 3). However, the morphological (gill-raker counts) and reproductive (winter instead of summer spawning) traits of these individuals did not match those of historical *C. gutturosus*. We therefore calculated pairwise F_{ST} with and without these genetically assigned *C. gutturosus*-like individuals. Pairwise genetic differentiation among the three extant species has dropped dramatically through time (Fig. 2b) and global F_{ST} has decreased over twofold (0.108/0.165 to 0.046/0.047, without/with *C. gutturosus*, respectively).

Speciation reversal should also increase genetic variation within extant species. Consistent with this prediction, allelic richness has increased through time in *Coregonus wartmanni* ($N = 10$, d.f. = 8, $t = 3.38$, $P = 0.009$) and a similar trend is seen in *Coregonus macrophthalmus* ($N = 10$, d.f. = 8, $t = 2.17$, $P = 0.062$; Supplementary Table 4). Out of 11 alleles found only in *C. gutturosus* among the pre-eutrophication samples (private alleles), 5 were found in contemporary *Coregonus* species of Lake Constance (Supplementary Table 5). The probability of finding at least one of these alleles in pre-eutrophication samples of the other species, assuming similar frequencies, is 98% and suggests that the extinction of *C. gutturosus* involved hybridization with other species.

Data from Lake Brienz, the lake that is least polluted and that has no loss in species or functional diversity (Table 1), contrast and complement those from Lake Constance. For the three endemic species, global genetic differentiation (global F_{ST}) was historically (1952–70) identical to that in Lake Constance (0.166) but has not declined until the present (0.183). Moreover, no significant increase in allelic richness was observed in any of the three species. Nevertheless, out of 12 historically private alleles of the summer- and deep-spawning *Coregonus albellus*, 7 were also found in contemporary samples of other species, suggesting that gene flow between species has also occurred in this lake (see also ref. 33).

Additional support for speciation reversal comes from Lakes Zürich and Walen, which share a single-origin species pair, a small, deep-spawning species (*Coregonus heglingus*) and a large, shallow-spawning species (*Coregonus duplex*¹⁸; Supplementary Fig. 5). Despite a common evolutionary history, pairwise F_{ST} between the species

Table 1 | Whitefish species and functional diversity in 17 pre-alpine lakes.

Lake (no.)	Species diversity and genetic differentiation					Functional ecological diversity					Mean egg survival (%)	Oxygenated lake depth (m)	P_{\max} ($\mu\text{g l}^{-1}$)	Maximum lake depth (m)
	Historical species	Contemporary species	Species loss (%)	Genetic species	Change in genetic differentiation (%)	Historical gill-raker range	Contemporary gill-raker range	Functional loss (%)	N Historical gill-raker range	N Contemporary gill-raker range				
Lake Geneva (1)	2	0 (1†)	−100			15	10*	−33.30	61	24*	84.4	254.17	90	309
Lake Neuchâtel (2)	2	1.5	−25	1.5		17	17	0.00	?	341	43.8	153.00	50	152
Lake Murten (3)	2	0	−100			17	12*	−29.40	?	30*		8.93	150	45.5
Lake Biel (4)	2	2	0			17	19	11.80	?	49	17.9	27.50	132	74
Lake Thun (5)	5	5	0	5		33	29	−12.10	471	331	67.2	214.00	21	217
Lake Brienz (6)	3	3	0	3	10	18	20	11.10	>123	100		243.97	17	261
Lake Sempach (7)	2	0	−100	1		13	13*	0.00	>12	76*	0.7	8.26	165	87
Lake Lucerne (8)	4	4	0	4		23	23	0.00	180	730	42.0	203.49	34	214
Lake Zug (9)	2	1	−50			21	11*	−47.62	?	20*	0.3	8.50	208	198
Lake Baldeg (10)	1	0	−100			17			?			4.34	517	66
Lake Hallwil (11)	1	0	−100			17	11*	−35.30	?	20*	0.9	6.69	260	47
Lake Zürich (12)	3	2	−33	2		18	18	0.00	76	66	35.3	9.72	119	136
Lake Walen (13)	3	2	−33	2		19	20	5.30	?	236	37.8	144.00	26	145
Lake Greifen (14)	1	0	−100			11	11*	0.00	?	50*		0.00	507	32.3
Lake Pfäffiker (15)	1	0	−100			15	9*	−40.00	?	19*		0.00	367	35
Lake Constance (16)	5	4	−20	3	−57 (−71.5†)	36	26*	−27.80	694	79*	31.4	248.91	87	254
Lake Sarnen (17)	2	1	−50			13	10*	−23.10	?	20*	59.4	47.33	21	52
Total/average	41	25.5	−38		−24 (−31†)			−13.78		2,191				

The number of historically observed and presently observed phenotypically distinct, naturally recruiting whitefish species (Historical species²² and Contemporary species, respectively). ‡The present wild population observed in Lake Geneva does not correspond to either of the two described species; the percentage loss in species numbers (Species loss); the number of genetically distinct species observed today (Genetic species), in which 1.5 represents a species cline observed in Lake Neuchâtel²³; the percentage reduction in global genetic differentiation; the historically and currently observed gill-raker count range (Historical gill-raker range and Contemporary gill-raker range, respectively); the functional diversity (Functional loss); the sample sizes for historical data (Historical gill-raker range, N) and contemporary gill-raker analysis (N Contemporary gill-raker range), the mean egg survival (Mean egg survival), the biologically available depth during eutrophication with more than 2.5 mg l^{−1} dissolved oxygen (Oxygenated lake depth); the maximum total phosphorus concentration observed during the eutrophic period (P_{\max}); and the maximum lake depth (Maximum lake depth).

* Gill-raker ranges adjusted for unequal sample sizes (see Methods, Supplementary Information, Supplementary Table 6 and Supplementary Fig. 4).

† Number in brackets corresponds to the loss if the phenotypically extinct *C. gutturosus* is included in the analysis (see Supplementary Table 3).

today in eutrophic Lake Zürich (0.041) is less than half of that in oligotrophic Lake Walen (0.110).

Phenotypic signs of speciation reversal

Speciation reversal is expected to erode interspecific phenotypic distinctiveness^{4,6}. Gill-raker counts provide a measure of heritable phenotypic trophic adaptation¹⁹, and the contemporary range in gill-raker number and total body shape disparity of individuals in a lake are correlated ($N = 15$, slope = 0.49; $R^2 = 0.36$, $P < 0.011$; Supplementary Fig. 6). Across lakes, the distances of species means from the historical midpoint of species means in a lake have become significantly smaller over time ($N = 19$, $t = 2.56$, d.f. = 18, $P = 0.020$). Extant species have converged in moderately and strongly polluted lakes ($N = 10$, $t = 2.43$, d.f. = 9, $P = 0.038$, Fig. 2e) but not in weakly and mildly polluted lakes ($N = 9$, $t = 1.06$, d.f. = 8, $P = 0.319$). Relative contemporary disparity (see phenotypic tests of speciation reversal in Supplementary Information) was significantly lower in moderately and strongly polluted lakes than in weakly and mildly polluted lakes ($N = 5$ (moderately and strongly polluted lakes), $N = 6$ (weakly and mildly polluted lakes), $t = 2.48$, d.f. = 9, $P = 0.035$; Fig. 2e). The best general linear model contained maximum phosphorus concentration, maximum lake depth and oxygenated depth, with phosphorus having the largest and most significant effect ($N = 10$ lakes, $R^2 = 0.85$, $P < 0.001$, $\Delta\text{AICc} = 7.95$; regression coefficient for phosphorus -0.79 , $P = 0.002$).

In all but two radiations, species with few gill-rakers spawn in shallow water, whereas species with many gill-rakers spawn deeper^{22,23}. Speciation reversal predicts that the range in gill-raker number should contract from both ends of the distribution, whereas extinction through demographic decline of deep spawners predicts a contraction at the high end of the distribution. Consistent with speciation reversal, diversity has been lost from both ends of the distribution in each lake that experienced a range reduction (Table 1), independent of whether the two deep-spawning species with a low gill-raker count were included or not (mean loss at lower end is -3.4 or -2.7 gill rakers, respectively; Wilcoxon test: $Z = -2.54$, $N = 8$, d.f. = 7, $P = 0.011$ for both cases; mean loss at upper end is -2.75 , $Z = -2.54$, $N = 8$, d.f. = 7, $P = 0.011$ for both cases). This result was robust to the removal of Lakes Murten, Hallwil and Pfäffiker where natural recruitment had

ceased and stocks are maintained by stocking from hatcheries (mean loss at lower end is -3.4 or -2.4 gill rakers, respectively; Wilcoxon test, $Z = -2.03$, $N = 5$, d.f. = 4, $P = 0.042$ for both cases; mean loss at upper end is -3 ; $Z = -2.03$, $N = 5$, d.f. = 4, $P = 0.010$ for both cases).

Thus, in cases in which several species persisted in sympatry after eutrophication, their phenotypes converged, and the extinction of species was associated with evolution to intermediate phenotypes in the remaining species. This is consistent with partial and complete speciation reversal, respectively.

Discussion

Our evidence suggests anthropogenic eutrophication has led to speciation reversal in whitefish radiations by increasing gene flow between previously ecologically differentiated species. Although divergent natural selection could in principle maintain species differences in the face of increased gene flow, eutrophication seems to have altered reproductive and ecological niche spaces to the degree that selection cannot counteract the homogenizing effects of gene flow. It is possible that accidental hybridization in hatcheries has contributed to interspecific gene flow. However, while reductions in genetic differentiation were related to eutrophication, hatcheries operate on all lakes. Thus, this alone cannot explain observed patterns of diversity loss.

The study lakes have lost 38% of species diversity, 14% of functional diversity and 28% of functional disparity among species. At least eight endemic species and seven distinct populations of extant species have become extinct (Table 1 and Supplementary Table 1). Only 4 of 17 lakes suffered no species loss. Among remaining species, genetic differentiation is reduced. This loss of species richness, phenotypic diversity and genetic differentiation occurred mainly unnoticed despite the commercial importance of whitefish. Similar large losses of whitefish diversity may have occurred in other lakes outside Switzerland (Supplementary Table 1) and the extinction of endemic char species pairs in some of the same lakes could have involved similar mechanisms²¹. Finally, we also note that similar patterns of diversity loss have been observed in several other taxa^{4–6,15,34,35}.

Loss of biodiversity through speciation reversal may be underappreciated for two reasons. First, the process can be difficult to detect because it does not require changes in distribution or abundance but can manifest through subtle changes in patterns of variation within

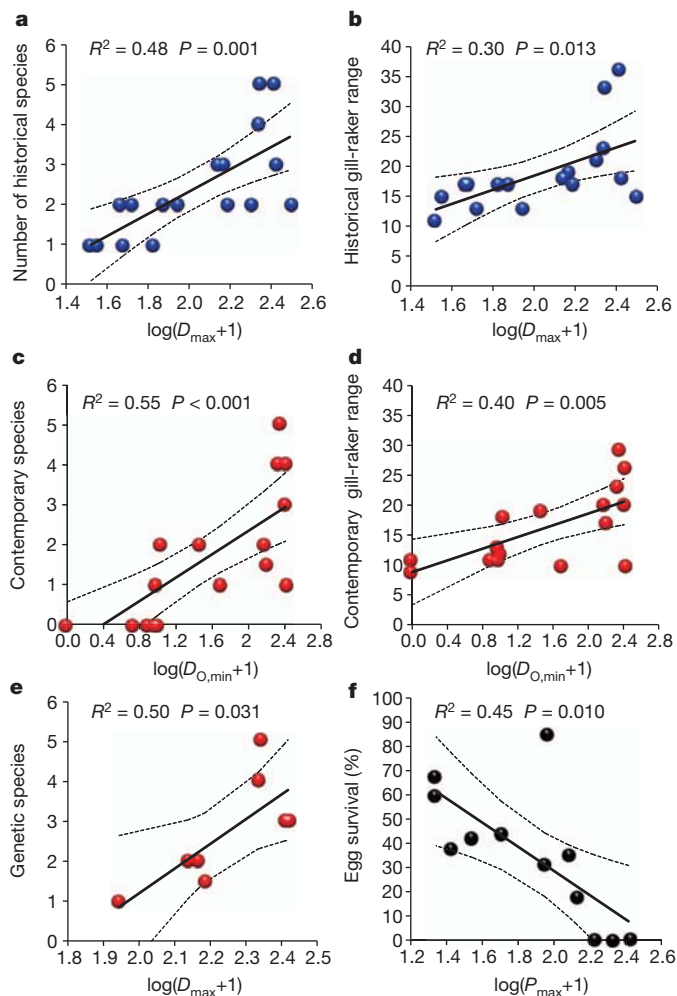


Figure 3 | Whitefish diversity explained by environmental variables. **a, b**, Historical whitefish diversity as species numbers (**a**) and the range in gill-raker numbers (**b**), plotted against maximum lake depth (D_{\max}). **c, d**, Contemporary diversity, measured as species numbers (**c**) and the range in gill-raker numbers (**d**), plotted against oxygenated depth ($D_{O,\min}$). **e**, The number of contemporary genetically differentiated species plotted against maximum lake depth. **f**, The relationship between the maximum phosphorus concentration (P_{\max}) and the viability of the whitefish eggs in 12 lakes. The dashed lines represent the 95% confidence intervals for the regression line.

multi-species assemblages⁴. Second, speciation reversal is a potentially rapid process, by which species can collapse in just a few generations^{5,6,14}. Compelling tests of speciation reversal will often require historic samples with DNA of sufficient quality. Our results add to a growing body of evidence suggesting that freshwater fish radiations, but also terrestrial radiations^{14,15}, are threatened by anthropogenic activities that disrupt the ecological conditions and evolutionary processes that promote adaptive radiation^{4–6,35}. There is evidence from lake ecosystems that eutrophication-mediated speciation reversal may threaten diversity simultaneously at interacting trophic levels³⁶, and the effects on food webs require investigation. If the loss of ecologically dominant species, such as planktivorous fish, affects other ecosystem components, the impacts of speciation reversal may extend beyond the simple loss of species^{37,38}. Regardless of the mechanistic details, preserving ecosystem services requires maintaining functional ecosystems, which in turn requires protecting the ecological conditions and evolutionary mechanisms that generate and maintain species diversity^{2,37,39,40}.

METHODS SUMMARY

Between 2004 and 2010 we collected 2,449 whitefish from 16 lakes. Muscle tissue was preserved in 100% ethanol for genetic analyses. The first gill arch was

removed from 2,191 individuals for gill-raker counts (Table 1). Scale samples were used to analyse historical trends in genetic differentiation of species in Lakes Constance (1926–50: $N = 133$; 1969–80: $N = 92$) and Brienz (1952–70: $N = 66$). We collected data on historical species richness in 17 lakes and on contemporary richness in 16 lakes. We determined three different metrics of historical diversity and four metrics of contemporary diversity for each lake assemblage: first, species richness, identified using morphology, spawning ecology and taxonomic literature; second, the observed range in gill-raker number, a measure of heritable functional diversity; third, genetic species differentiation, using genotypes based on ten micro-satellite loci (for methods see ref. 23); fourth, phenotypic distinctiveness of species using gill-raker mean counts. When possible, individual assignments to species was based on a Bayesian population inference algorithm (STRUCTURE version 2.3.3⁴¹; 30,000 burn in and 300,000 Markov chain Monte Carlo steps). Environmental variables for each lake were obtained from the literature⁴² and government databases. Maximum phosphorus concentration corresponds to the highest value observed between 1951 and 2004. Oxygenated depth was the minimum depth range observed during the eutrophic phase with the water containing at least 2.5 mg l^{-1} dissolved oxygen (see environmental variables in Supplementary Information). Whitefish eggs were collected from the lake bottom in twelve lakes on several samplings between 1968 and 2008, and the percentage of normally developing eggs was calculated⁴³.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 5 February 2011; accepted 3 January 2012.

- Chapin, F. S. *et al.* Consequences of changing biodiversity. *Nature* **405**, 234–242 (2000).
- Rosenzweig, M. L. Loss of speciation rate will impoverish future diversity. *Proc. Natl Acad. Sci. USA* **98**, 5404–5410 (2001).
- Rhymer, J. M. & Simberloff, D. Extinction by hybridization and introgression. *Annu. Rev. Ecol. Syst.* **27**, 83–109 (1996).
- Seehausen, O. Losing biodiversity by reverse speciation. *Curr. Biol.* **16**, R334–R337 (2006).
- Seehausen, O., Van Alphen, J. J. M. & Witte, F. Cichlid fish diversity threatened by eutrophication that curbs sexual selection. *Science* **277**, 1808–1811 (1997).
- Taylor, E. B. *et al.* Speciation in reverse: morphological and genetic evidence of the collapse of a three-spined stickleback (*Gasterosteus aculeatus*) species pair. *Mol. Ecol.* **15**, 343–355 (2006).
- Grant, B. R. & Grant, P. R. Fission and fusion of Darwin's finches populations. *Proc. R. Soc. B* **363**, 2821–2829 (2008).
- Gilman, R. T. & Behm, J. E. Hybridization, species collapse, and species reemergence after disturbance to premating mechanisms of reproductive isolation. *Evolution* **65**, 2592–2605 (2011).
- Schluter, D. *The Ecology of Adaptive Radiation* (Oxford Univ. Press, 2000).
- Coyne, J. A. & Orr, H. A. *Speciation* (Sinauer Associates, 2004).
- Rundle, H. D. & Nosil, P. Ecological speciation. *Ecol. Lett.* **8**, 336–352 (2005).
- Schluter, D. Evidence for ecological speciation and its alternative. *Science* **323**, 737–741 (2009).
- Servedio, M. R. *et al.* Magic traits in speciation: ‘magic’ but not rare? *Trends Ecol. Evol.* **26**, 389–397 (2011).
- Hendry, A. P. *et al.* Possible human impacts on adaptive radiation: beak size bimodality in Darwin's finches. *Proc. R. Soc. B* **273**, 1887–1894 (2006).
- De León, L. F. *et al.* Exploring possible human influences on the evolution of Darwin's finches. *Evolution* **65**, 2258–2272 (2011).
- Schluter, D. Ecological speciation in postglacial fishes. *Proc. R. Soc. B* **351**, 807–814 (1996).
- Rundle, H. D., Nagel, L., Boughman, J. W. & Schluter, D. Natural selection and parallel speciation in sympatric sticklebacks. *Science* **287**, 306–308 (2000).
- Hudson, A. G., Vonlanthen, P. & Seehausen, O. Rapid parallel adaptive radiations from a single hybridogenic ancestral population. *Proc. R. Soc. B* **278**, 58–66 (2011).
- Bernatchez, L. in *Evolution Illuminated* (eds Hendry, A. P. & Stearns, S. C.) 175–207 (Oxford Univ. Press, 2004).
- McPhail, J. D. Ecology and evolution of sympatric sticklebacks (*Gasterosteus*)—origin of the species pairs. *Can. J. Zool.* **71**, 515–523 (1993).
- Kottelat, M. & Freyhof, J. *Handbook of European Freshwater Fishes* (Kottelat, Cornol and Freyhof, 2007).
- Steinmann, P. Monographie der schweizerischen koregonen. Beitrag zum problem der entstehung neuer arten. Spezieller teil. *Schweiz. Z. Hydrobiol.* **12**, 340–491 (1950).
- Vonlanthen, P. *et al.* Divergence along a steep ecological gradient in lake whitefish (*Coregonus* sp.). *J. Evol. Biol.* **22**, 498–514 (2009).
- Woods, P. J., Müller, R. & Seehausen, O. Intergenomic epistasis causes asynchronous hatch times in whitefish hybrids, but only when parental ecotypes differ. *J. Evol. Biol.* **22**, 2305–2319 (2009).
- Müller, R. & Stadelmann, P. Fish habitat requirements as the basis for rehabilitation of eutrophic lakes by oxygenation. *Fish. Mgmt. Ecol.* **11**, 251–260 (2004).
- Verschuren, D. *et al.* History and timing of human impact on Lake Victoria, East Africa. *Proc. R. Soc. B* **269**, 289–294 (2002).
- Smith, V. H. & Schindler, D. W. Eutrophication science: where do we go from here? *Trends Ecol. Evol.* **24**, 201–207 (2009).

28. Straile, D. & Geller, W. The response of *Daphnia* to changes in trophic status and weather patterns: a case study from Lake Constance. *ICES J. Mar. Sci.* **55**, 775–782 (1998).
29. Jeppesen, E., Jensen, J. P., Søndergaard, M., Lauridsen, T. & Landkildehus, F. Trophic structure, species richness and biodiversity in Danish lakes: changes along a phosphorus gradient. *Freshwat. Biol.* **45**, 201–218 (2000).
30. Blumenshine, S. C., Vadeboncoeur, Y., Lodge, D. M., Cottingham, K. L. & Knight, S. E. Benthic-pelagic links: responses of benthos to water-column nutrient enrichment. *J. N. Am. Benthol. Soc.* **16**, 466–479 (1997).
31. Powers, S. P. *et al.* Effects of eutrophication on bottom habitat and prey resources of demersal fishes. *Mar. Ecol. Prog. Ser.* **302**, 233–243 (2005).
32. Waples, R. S. & Do, C. Linkage disequilibrium estimates of contemporary N_e using highly variable genetic markers: a largely untapped resource for applied conservation and evolution. *Evol. Appl.* **3**, 244–262 (2010).
33. Bittner, D., Excoffier, L. & Largiadere, C. R. Patterns of morphological changes and hybridization between sympatric whitefish morphs (*Coregonus* spp.) in a Swiss lake: a role for eutrophication? *Mol. Ecol.* **19**, 2152–2167 (2010).
34. Seehausen, O. *et al.* Speciation through sensory drive in cichlid fish. *Nature* **455**, 620–623 (2008).
35. Heath, D., Bettles, C. M. & Roff, D. Environmental factors associated with reproductive barrier breakdown in sympatric trout populations on Vancouver Island. *Evol. Appl.* **3**, 77–90 (2010).
36. Brede, N. *et al.* The impact of human-made ecological changes on the genetic architecture of *Daphnia* species. *Proc. Natl Acad. Sci. USA* **106**, 4758–4763 (2009).
37. Harmon, L. J. *et al.* Evolutionary diversification in stickleback affects ecosystem functioning. *Nature* **458**, 1167–1170 (2009).
38. Goldschmidt, T., Witte, F. & Wanink, J. Cascading effects of the introduced Nile perch on the detritivorous phytoplanktivorous species in the sublittoral areas of Lake Victoria. *Conserv. Biol.* **7**, 686–700 (1993).
39. Seehausen, O. Speciation affects ecosystems. *Nature* **458**, 1122–1123 (2009).
40. Faith, D. P. *et al.* Ecosystem services: an evolutionary perspective on the links between biodiversity and human well-being. *Curr. Opin. Env. Sust.* **2**, 1–9 (2010).
41. Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959 (2000).
42. Liechti, P. *Der Zustand der Seen in der Schweiz* (Schriftenreihe Umwelt Nr. 237; Bundesamt für Umwelt, Wald und Landschaft, 1994).
43. Müller, R. Trophic state and its implications for natural reproduction of salmonid fish. *Hydrobiologia* **243**, 261–268 (1992).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank all professional fishermen who provided fish specimens. We thank M. Kugler from the Amt für Natur, Jagd und Fischerei, St. Gallen and the institute of Seenforschung und Fischereiwesen Langenargen for providing historical whitefish scales from Lake Constance. We acknowledge the Swiss Federal Institute for Aquatic Science and Technology (EAWAG), the Internationale Gewässerschutzkommission für den Bodensee (IGKB) and the Federal Office for Environment (FOEN) for providing environmental data. We also thank G. Périat, S. Mwaiko, M. Barluenga, H. Araki, M. Maan, J. Brodersen, P. Nosil, K. Wagner and all members of the Fish Ecology and Evolution laboratory for assistance in the laboratory, and for comments and suggestions on the manuscript, B. Müller for help with the analysis of the oxygen profiles, and C. Melian for help with data analyses. We acknowledge financial support by the Eawag Action Field Grant ‘AquaDiverse—understanding and predicting changes in aquatic biodiversity’ (to O.S.).

Author Contributions P.V. contributed to conception and design of the study, collected fish, generated gill-raker and contemporary genetic data, and carried out most of the statistical analyses. D.B. collected fish, generated gill-raker, historical and contemporary genetic data. A.G.H. collected fish and generated gill-raker and geometric morphometric data. K.A.Y. participated in designing the study and writing the manuscript. R. M. collected and analysed egg data and contributed to fish collection. B.L.-H. contributed to fish, gill-raker, and genetic data collection. D.R. contributed to analyses and writing. S.D.P. contributed to collection of historical genetic data. C.R.L. supervised parts of sampling, gill-raker counting and contemporary genetic data collection. O.S. conceived and designed the project, supervised the project, and contributed to data analyses. P.V. and O.S. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to O.S. (ole.seehausen@eawag.ch).

METHODS

Sampling. Between 2004 and 2010 we collected 2,449 whitefish from 16 lakes. We collected at least 20 individuals of each known species from 16 lakes, except *C. heglungus* in Lake Zürich (17 individuals), and *C. sp. 'Felchen'* in Lake Thun, which could not be obtained. In most lakes, we collected fish directly on the spawning grounds. In six lakes (Lakes Sempach, Walen, Lucerne, Thun, Brienz and Neuchâtel), fish were collected several times from many different spawning sites to distinguish intraspecific genetic population structure and species structure. No geographical or temporal differences within species could be observed (ref. 23 and B.L.H., personal communication). We sampled systematically along water-depth gradients during the spawning period in Lakes Neuchâtel²³ and Lucerne (B.L.-H., P.V., A.G.H., K. Lucek and O.S., unpublished data). The length, weight and sex of every fish was recorded. Muscle tissue was removed and preserved in 100% ethanol for DNA analysis. The first gill arch was removed from 2,191 individuals for gill-raker counting (Table 1). Scale samples were used for molecular genetic analyses of historical trends in species differentiation in Lake Constance (1926–50: $N = 133$; 1969–80: $N = 92$) and Lake Brienz (1952–70: $N = 66$).

Historical and contemporary diversity. We collected data on historical and contemporary diversity in 17 and 16 Swiss lakes, respectively. We determined three different metrics of historical and four of contemporary diversity (details in Supplementary Information). First, contemporary species richness was determined using the same traits and procedures as Steinmann in 1950, who determined historical species richness using morphological and meristic traits, and information on spawning ecology²²; Second, contemporary ranges in gill-raker numbers were collected from our recent samples (see above) and historical gill-raker data were taken from Steinmann²². In whitefish, gill-raker number is related to feeding ecology⁴⁴ and is highly heritable (0.79)¹⁹, and thus provides an ecologically meaningful and taxonomically independent (Supplementary Fig. 7) estimate of heritable functional diversity. To enable comparisons between historical and contemporary data when sample sizes were unequal or (for historical data) unknown, we used available data for each species to create normal distributions from which 100 virtual individuals were then randomly sampled. Third, genetic species differentiation was determined by genotyping historical and contemporary samples at 10 microsatellite loci. Details of laboratory methods that were used for contemporary samples are given in ref. 23. Whenever possible, the identification of sympatric genetically differentiated species and individual assignment were performed using the Bayesian population inference algorithm in STRUCTURE version 2.3.3 (ref. 41) (30,000 burn in and 300,000 MCMC steps). However, STRUCTURE is typically inefficient when $F_{ST} < 0.05$ (unless many loci are sampled)⁴⁵. This was found to be the case between some species in Lake Lucerne and in Lake Zürich. A combination of morphology and spawning ecology was used to identify species in these lakes that was confirmed a posteriori by significant F_{ST} values observed between species sampled in sympatry. We calculated the extent of contemporary genetic differentiation among species in eight lakes, and the historical differentiation in two lakes; one that was moderately impacted (Lake Constance) and the other little impacted (Lake Brienz) by eutrophication. Fourth, phenotypic distinctiveness of species was determined using gill-raker mean counts for each species in each lake.

DNA extraction and PCR amplification of DNA from historical material. Total DNA was extracted from historical dried scales using a modified standard phenol-chloroform-ethanol extraction method⁴⁶. The DNA quantity was measured using a Nanodrop ND-1000 (Nanodrop technologies) spectrophotometer and all samples containing less than 20 ng μL^{-1} DNA were excluded from further investigations. Polymerase chain reaction (PCR) was performed according to the QIAGEN Multiplex standard protocol with an annealing temperature (T_{AN}) of 57 °C and 35 cycles (Sets 1 and 3) or 45 cycles (Sets 2 and 4). Denatured fragments

were resolved on an automated DNA sequencer (ABI 3100). Genotypes were determined with the GENEMAPPER 4.0 (ABI) software and checked visually. Each sample was amplified twice in a separate PCR. When both genotypes were identical we used these genotypes for further analysis (81.5% of all genotypes). When both genotypes were missing, no further attempt was taken to genotype a sample at that locus (8.9% of all genotypes). Finally, when only one of the two genotypes could be determined (9.6%), a third—or when needed, a fourth—separate PCR was performed to confirm genotypes. To estimate reproducibility, 28 samples were independently extracted and the procedure that is described above was repeated. 240 genotypes were compared and 8 mismatches were found (reproducibility, 96.7%). Only individuals with a minimum of six successfully genotyped loci were considered for population genetic data analysis. The level of missing data in loci with large fragment lengths was considerable in historical populations (40.5% for locus CoCl-61, 26.4% for locus CoCl-10 and 14.5% for locus CoCl-45). Separate analyses excluding these loci yielded very similar results (data not shown). Therefore, all analyses were performed including all loci.

Environmental variables. Lake depths (m) and maximum phosphorus content (P_{tot} (mg m^{-3})) data were obtained from ref. 42. O_2 depth profiles (mg l^{-1}) (Supplementary Table 7) were obtained from the Federal Office for the Environment (FOEN), Swiss Federal Institute of Aquatic Science and Technology (EAWAG) and the Internationale Gewässerschutzkommission für den Bodensee (IGKB). For maximum phosphorus concentration, we took the highest value that was observed in time series covering the period from 1951 to 2004, which includes the onset and peak of the eutrophic phase and the re-oligotrophication that began in the 1980s. The maximum depth of a lake was the depth measured from the lake surface to the deepest point of the lake. Oxygenated depth was calculated as the minimum water depth range observed during the eutrophic phase with the water containing at least 2.5 mg l^{-1} dissolved oxygen. The limit of 2.5 mg l^{-1} was chosen to correspond to the critical oxygen level at which embryo development is negatively affected⁴⁷.

Egg survival data. Whitefish eggs were collected from the lake bottom in 12 lakes on several occasions between 1968 and 2008. Sampling was done in each lake between early January and early March, before the anticipated beginning of mass hatching of the corresponding whitefish species. As a comparative measure of egg viability, we calculated the percentage of eggs that developed normally. Details of the sampling methods can be found in ref. 43 and in Supplementary Information.

Statistical analyses. We used least squares regressions and an information theoretic approach to select the models that best explain the relationship between predictor and response variables (Supplementary Information). Before comparing data, we tested for significant deviations from normal distributions using a Shapiro Wilks test. For data that did not significantly deviate from normality, a standard or paired Welch's t -test was used. When the data significantly deviated from normality, a Wilcoxon signed rank or a Mann–Whitney U test was used. All tests were two-tailed.

44. Harrod, C., Mallela, J. & Kahilainen, K. K. Phenotype-environment correlations in a putative whitefish adaptive radiation. *J. Anim. Ecol.* **79**, 1057–1068 (2010).
45. Latch, E. K., Dharmarajan, G., Glaubitz, J. C. & Rhodes, O. E. Relative performance of Bayesian clustering software for inferring population substructure and individual assignment at low levels of population differentiation. *Conserv. Genet.* **7**, 295–302 (2006).
46. Wasko, A. P., Martins, C., Oliveira, C. & Foresti, F. Non-destructive genetic sampling in fish. An improved method for DNA extraction from fish fins and scales. *Hereditas* **138**, 161–165 (2003).
47. Czerkies, P., Kordalski, K., Golas, T., Kryszinski, D. & Luczynski, M. Oxygen requirements of whitefish and vendace (Coregoninae) embryos at final stages of their development. *Aquaculture* **211**, 375–385 (2002).

Prions are a common mechanism for phenotypic inheritance in wild yeasts

Randal Halfmann^{1,2†*}, Daniel F. Jarosz^{1*}, Sandra K. Jones^{1†}, Amelia Chang^{1,2†}, Alex K. Lancaster¹ & Susan Lindquist^{1,2,3}

The self-templating conformations of yeast prion proteins act as epigenetic elements of inheritance. Yeast prions might provide a mechanism for generating heritable phenotypic diversity that promotes survival in fluctuating environments and the evolution of new traits. However, this hypothesis is highly controversial. Prions that create new traits have not been found in wild strains, leading to the perception that they are rare ‘diseases’ of laboratory cultivation. Here we biochemically test approximately 700 wild strains of *Saccharomyces* for $[PSI^+]$ or $[MOT3^+]$, and find these prions in many. They conferred diverse phenotypes that were frequently beneficial under selective conditions. Simple meiotic re-assortment of the variation harboured within a strain readily fixed one such trait, making it robust and prion-independent. Finally, we genetically screened for unknown prion elements. Fully one-third of wild strains harboured them. These, too, created diverse, often beneficial phenotypes. Thus, prions broadly govern heritable traits in nature, in a manner that could profoundly expand adaptive opportunities.

The heritable variation that drives new forms and functions is generally ascribed to mutations in the genetic code. We previously proposed an entirely different pathway for creating heritable phenotypic diversity¹, through which the inheritance of new traits can precede the genetic changes that ultimately hardwire them. The mechanism for this seemingly paradoxical flow of information resides in epigenetic switches encoded entirely by self-perpetuating changes in protein structure, known as prions.

The best studied prion is the yeast translation-termination factor Sup35. Like other prions, Sup35 carries a prion-determining domain (PrD) that is dispensable for the protein's normal function. This PrD occasionally adopts an amyloid conformation. When it does, that amyloid perpetuates itself by templating the same conformation to the PrDs of other Sup35 molecules. This sequesters most Sup35 into insoluble fibres². The ensuing reduction in translation-termination activity increases stop codon read-through, producing a variety of new traits that depend upon previously cryptic genetic variation.

Just as the mitotic apparatus ensures inheritance of chromosomally determined traits, the prion-partitioning function provided by Hsp104 (refs 3, 4) ensures inheritance of prion phenotypes. Hsp104 is a molecular machine that severs prion fibres, allowing replicating prion templates to be faithfully inherited by daughter cells. The prion formed by Sup35 is known as $[PSI^+]$, brackets denoting its cytoplasmic inheritance and capital letters its dominant phenotypes.

Cells expressing Sup35 in the non-prion $[psi^-]$ state spontaneously switch to $[PSI^+]$ at a frequency of about 1 in 10^6 (refs 5, 6). We have proposed that $[PSI^+]$ provides a beneficial ‘bet-hedging’ mechanism to enhance survival in the face of fluctuating environments: by the time a yeast colony has reached appreciable size, a few $[PSI^+]$ cells will have appeared, expressing heritable new traits. If the trait is detrimental, only a few individuals in a large population will be lost. However, if it is advantageous, those few cells might ensure survival under conditions when the population would otherwise perish. $[PSI^+]$ is also lost sporadically. This guarantees that $[psi^-]$ cells

will arise in $[PSI^+]$ colonies, providing a complementary survival advantage.

A particularly attractive feature of this mechanism is that it provides immediate access to genetically complex traits^{1,7}. Regions downstream of stop codons frequently accumulate genetic variation. $[PSI^+]$ -mediated read-through allows this previously cryptic variation to have biological consequences at multiple loci simultaneously. The complex traits produced by this prion would be less likely to evolve if the individually contributing mutations had to be selected for as they arose. In the long run, reduced translational fidelity should be detrimental. However, advantageous phenotypes initially dependent on $[PSI^+]$ might be assimilated by various means⁷, allowing the prion to be lost and the trait maintained.

Several lines of evidence support this hypothesis. First, mathematically, even an infrequent selective advantage for $[PSI^+]$ would be sufficient to maintain Sup35's prion switching capacity^{8,9}. Second, the sequence of Sup35's PrD is highly divergent but has retained, for at least 500 million years, two unusual features that regulate bi-stable inheritance of prion and non-prion phenotypes. An extreme amino acid bias in one segment drives the PrD into a self-templating prion amyloid, whereas an immediately adjacent charged segment stabilizes it in the soluble non-prion state. Third, the rates at which cells switch into and out of the prion state increases when cells are not well-suited to their environments and new phenotypes have a better chance of being beneficial¹⁰. Increased switching is a direct consequence of the effects that diverse environmental stresses have on protein folding and homeostasis^{11,12} and also fulfils a critical theoretical prediction for such an evolvability function^{6,13}.

In addition to Sup35, at least two dozen other proteins can form prions that are transmitted through the prion-partitioning activity of Hsp104 (refs 14, 15) in laboratory yeast. These prions are strikingly enriched in transcription factors and RNA-processing proteins that are well situated to transduce genetic variation into phenotypic effect. They too, therefore, might enable the inheritance of diverse biological

¹Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA. ²Department of Biology, MIT, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA. ³Howard Hughes Medical Institute, MIT, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA. [†]Present addresses: University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, Texas 75390-9038, USA (R.H.); The Rockefeller University 1230 York Avenue, New York, New York 10065, USA (S.K.J.); Harvard Medical School, 25 Shattuck Street Boston, Massachusetts 02115, USA (A.C.).

*These authors contributed equally to this work.

traits, enhancing survival in fluctuating environments. However, as attractive as such ideas may be¹⁶, and as intensely studied as yeast prions have been, their proposed adaptive value remains highly controversial^{17–19}. Indeed, prions are often categorized as rare ‘diseases’ of yeast or mere artefacts of laboratory cultivation. A key argument is that $[PSI^+]$ and other prions with phenotypic consequences have not been found in wild strains, despite attempts to find them. Here we establish the natural biological importance of prions through biochemical and biological analyses of hundreds of wild strains.

$[PSI^+]$ occurs in wild strains

To search for $[PSI^+]$ in wild strains we took advantage of the unusual stability of prion amyloid assemblies in ionic detergents, which enables their identification by semi-denaturing detergent–agarose gel electrophoresis (SDD–AGE)⁴. We modified the technique to enable high-throughput detection. Ultimately we analysed 690 yeast strains from diverse ecological niches (Supplementary Table 1). Amyloid polymers of Sup35 were present in ten (Fig. 1a, Supplementary Table 1 and Supplementary Fig. 1).

To ensure that these strains were not simply derived from a recent, prion-containing common ancestor, we sequenced the genomes of two. Over 25,000 single nucleotide polymorphisms established their independent origins (Supplementary Fig. 2). We also sequenced the *SUP35* gene in several of the strains, which established that they, too, had independent origins (Supplementary Table 2).

Do the Sup35 amyloids in these strains represent true prions? Owing to its central role in the inheritance of prion templates, even transient inhibition of Hsp104’s protein remodelling activity heritably ‘cures’ cells of their prion elements. We inhibited Hsp104 function by growth on medium containing low concentrations of guanidine hydrochloride (GdHCl), which selectively inhibits its ATPase activity, and then plated cells back to media without GdHCl. In all cases this

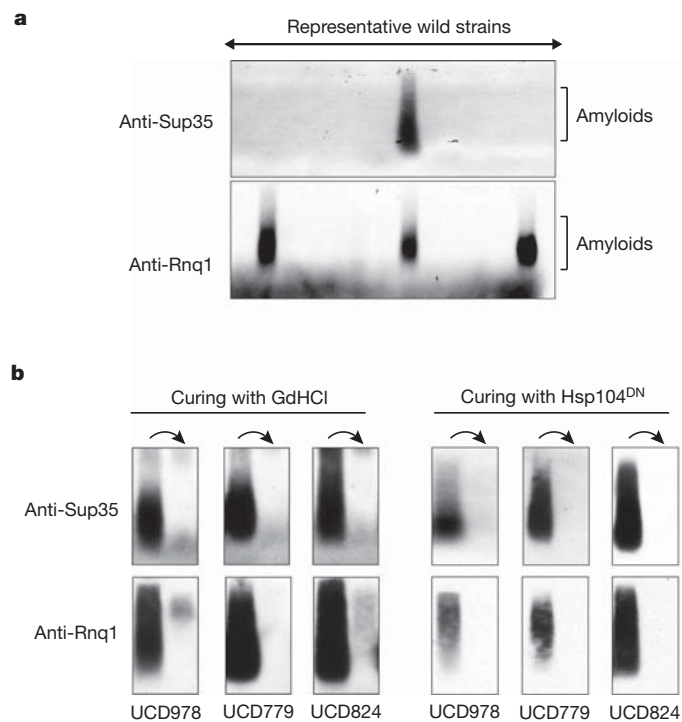


Figure 1 | Identification and verification of prions in wild yeast.

a, Representative SDD–AGE blot of wild yeast isolates probed with antibodies against Sup35 and Rnq1. Amyloid polymers produce characteristic smears. SDD–AGE does not reliably detect monomeric proteins. **b**, Transient inactivation of Hsp104 by GdHCl or expression of a dominant negative mutant of Hsp104 eliminates these amyloids, indicating that they are $[PSI^+]$ and $[RNQ^+]$ prions.

eliminated the amyloid (black arrows, Fig. 1b and Supplementary Fig. 1B). To ensure that curing was not due to an off-target effect of GdHCl we used a genetic approach—transiently expressing a dominant negative Hsp104 variant³, Hsp104^{DN}, on a plasmid marked with antibiotic resistance. (Wild strains contain no auxotrophies.) This also cured cells of the amyloid, confirming their prion-based inheritance (black arrows, Fig. 1b and Supplementary Fig. 1b, c).

Laboratory culture is not responsible

Might these prions have arisen simply as an artefact of laboratory culture conditions? In archiving wild strains, great care is taken to maintain their wild character (personal communication, L. Bisson). To determine directly if the conditions used might have inadvertently selected for the *de novo* appearance of $[PSI^+]$, we compared growth of the archived strains with their cured derivatives on all of the relevant media. No growth advantage was found for $[PSI^+]$ on any of these media (yeast potato dextrose, YPD, YM broth, FM broth, wort agar or Wallerstein nutrient agar) in any of the strains, and it was sometimes detrimental (Supplementary Table 3). However, in strain 5672 $[PSI^+]$ produced an extreme selective advantage on synthetic grape must, a medium that recapitulates conditions of early wine fermentation²⁰. This suggests the prion was advantageous in the strain’s natural ecological niche (Supplementary Fig. 3). In any case, this and several other experiments (Supplementary Information), indicated that prions harboured by the wild strains almost certainly originated in the yeasts’ natural environments.

Wild $[PSI^+]$ is associated with $[RNQ^+]$

In the laboratory, Sup35’s switch to the prion state strongly depends on the prion-enabling factor $[RNQ^+]$ ²¹. $[RNQ^+]$ is itself a prion formed by the Rnq1 protein²². $[RNQ^+]$ is the only prion previously known to exist in wild strains^{18,23,24}. We screened our collection for $[RNQ^+]$ amyloids, finding them in 43 strains (Fig. 1b). These, too, depended on the prion-partitioning factor Hsp104 (arrows, Fig. 1b and Supplementary Fig. 1). The correlation between $[RNQ^+]$ and $[PSI^+]$ ($P < 0.0001$, Fisher’s exact test) was striking: all the $[PSI^+]$ strains contained $[RNQ^+]$. This strongly indicates that $[RNQ^+]$ acts as a prion-inducing factor for $[PSI^+]$ in nature.

$[PSI^+]$ transforms natural genetic variation

Do wild prions generate phenotypic diversity from otherwise-cryptic natural variation? We compared growth of the wild $[PSI^+]$ strains with that of their cured derivatives in four carbon sources, under osmotic, oxidative, pH or ethanol stresses, and in the presence of antifungal drugs or DNA damaging agents. We also assessed their ability to invade the growth substratum.

Prion curing produced many phenotypic changes that varied with the genetic background (Fig. 2a). For example, strain UCD824, isolated from white wine, was resistant to acidic conditions and to fluconazole. UCD939, isolated from Lambrusco grapes, was resistant to the DNA-damaging agent 4-nitroquinoline 1-oxide (4-NQO). These beneficial phenotypes were greatly reduced by prion curing. UCD978, isolated from Beaujolais wine, was sensitive to the oxidative stressor tBOOH and became more resistant on curing. This same strain normally penetrated the agar surface, but this ability was lost after prion curing (Fig. 3a). Thus, in UCD978 the prion produced a trade-off, creating traits that were potentially detrimental or beneficial, depending on the circumstances.

GdHCl and Hsp104^{DN} cures cells of other Hsp104-dependent prions in addition to $[PSI^+]$ ^{14,15}. To determine if such curable phenotypes arose from $[PSI^+]$ itself, we transformed the ten strains with a plasmid expressing a Sup35 variant lacking the PrD (Sup35ΔPrD). This protein is immune to $[PSI^+]$ -mediated sequestration and restores normal translation termination in $[PSI^+]$ cells without altering other prions¹⁴. In most cases Sup35ΔPrD produced the same changes as curing with GdHCl and Hsp104^{DN} (Fig. 2b, Fig. 3b, see Supplementary

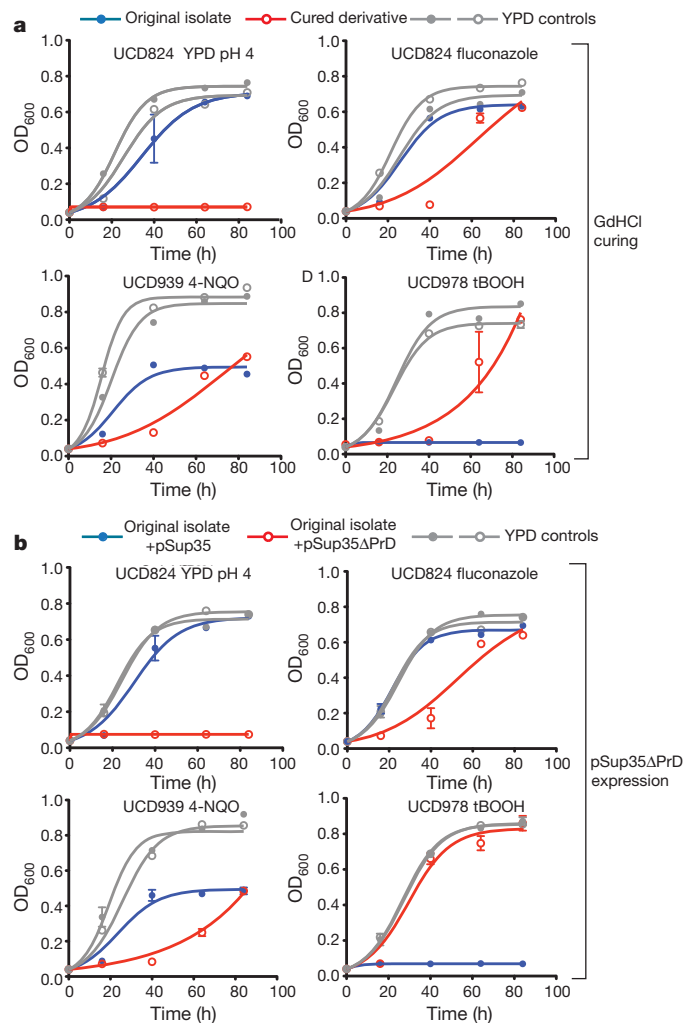


Figure 2 | Prion-contingent phenotypes of $[PSI^+]$ isolates. **a**, Wild $[PSI^+]$ strains show diverse phenotypes that are eliminated by transient Hsp104 inhibition. Growth curves for wild strains and cured derivatives in the indicated selective condition are in closed blue circles and open red circles, respectively. Growth in YPD is presented for each wild strain (closed grey circles) and its cured derivative (open grey circles) as a control. OD₆₀₀, optical density measured at 600 nm. **b**, Phenotypes of the wild $[PSI^+]$ strains were also eliminated by expression of Sup35ΔPrD. Growth curves for wild strains expressing Sup35 or Sup35ΔPrD in the indicated selective condition are in closed blue circles and open red circles, respectively. Growth in YPD for the indicated wild strain expressing Sup35 (closed grey circles) or Sup35ΔPrD (open grey circles) is presented as a control. Error bars are present on all points and represent the standard deviation from four independent biological replicates.

Information for discussion). Thus, most of the original traits were due to $[PSI^+]$.

Fixation of a $[PSI^+]$ -dependent phenotype

When laboratory strains of diverse backgrounds are crossed and sporulated, meiotic re-assortment of the genetic variants they contain can lead to the fixation of a prion trait^{1,7}. That is, whereas the trait initially depends on the prion, it can become prion-independent. Might this mechanism allow wild strains to drive prion-dependent traits to fixation? Wild yeasts frequently harbour considerable heterozygosity, and sequencing has revealed that the $[PSI^+]$ strain UCD978 was highly polymorphic. We asked if simple re-assortment of these polymorphisms could fix a $[PSI^+]$ -dependent trait.

Thirty random haploid progeny of UCD978 were tested for agar adhesion before and after curing. Five retained $[PSI^+]$ -dependent adhesion; twenty were no longer adhesive, with or without $[PSI^+]$;

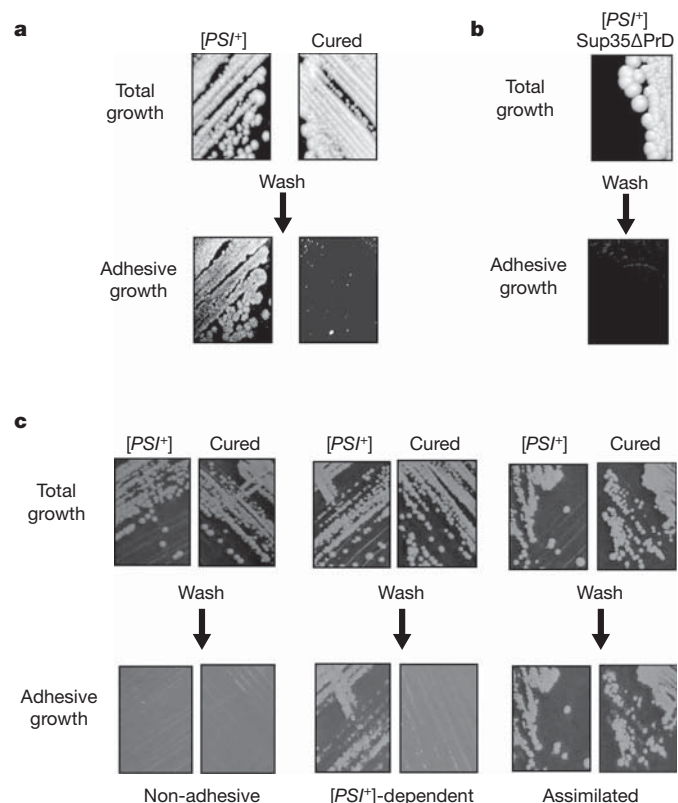


Figure 3 | Genetic assimilation of the $[PSI^+]$ -dependent adhesive phenotype in meiotic progeny of UCD978. **a**, $[PSI^+]$ allows the wine yeast UCD978 to adhere to agar surfaces. **a**, **b**, Adhesion is eliminated by GdHCl (**a**) or by expression of a non-aggregating version of Sup35, Sup35ΔPrD (**b**). **c**, Meiotic progeny of strain UCD978 show a diversity of phenotypes. In some spores the adhesive phenotype was assimilated and remained even after the $[PSI^+]$ prion was cured. In others the adhesive phenotype retained $[PSI^+]$ dependence. Finally, some lost the phenotype altogether, irrespective of $[PSI^+]$ status.

five remained adhesive even after $[PSI^+]$ curing (Fig. 3c). Given the frequency of fixation, it probably required the re-assortment of a few different polymorphisms. But clearly, the naturally occurring genetic variation present in this strain was alone sufficient to fix this trait.

$[MOT3^+]$ occurs in wild strains

Nearly two-dozen proteins with prion-forming capacity have recently been discovered in yeast (reviewed in ref. 15). Serendipitously, an endogenous hexahistidine motif in one, the transcriptional repressor Mot3, permits detection on SDD-AGE immunoblots¹⁴. Sixteen yeast proteins contain a hexahistidine motif, but only Mot3 has a prion-like sequence¹⁴. We found $[MOT3^+]$ amyloids in six of the 96 diverse strains we tested (Supplementary Table 1).

To determine if wild $[MOT3^+]$ prions produced potentially adaptive phenotypes, we first took advantage of Mot3's known function as a transcriptional repressor of genes involved in cell wall production. We tested wild $[MOT3^+]$ strains for resistance to the cell wall toxin calcofluor white. Strain Y-35, isolated from holly berries, was highly resistant to calcofluor. Resistance was heritably reduced by GdHCl treatment, and this treatment also eliminated $[MOT3^+]$ amyloids (Fig. 4a).

As a transcriptional repressor, when Mot3 switches into or out of its prion form it has the potential to broadly transform information flow. We next screened wild $[MOT3^+]$ strains and their cured derivatives against the same growth conditions used for the wild $[PSI^+]$ strains. Many phenotypes were altered by prion curing. For example, $[MOT3^+]$ NCYC 3311, a Finnish soil isolate, was resistant to acidic conditions. $[MOT3^+]$ Y-1537, isolated from grape must, was resistant

to fluconazole. Both traits were virtually eliminated by curing with GdHCl (Fig. 4b).

To determine if the traits were $[MOT3^+]$ -dependent, we expressed a Mot3 protein lacking the PrD (Mot3 Δ PrD) that is immune to prion sequestration but retains normal transcriptional function (R.H., personal communication). Analogous to Sup35 Δ PrD, this eliminates $[MOT3^+]$ phenotypes without affecting other prions. NCYC 3311 lost acid resistance and Y-1537 lost fluconazole resistance with this plasmid, but not with plasmids expressing the full-length protein (Fig. 4b). These phenotypes were, therefore, $[MOT3^+]$ -dependent. More broadly, the divergent consequences of this prion in different strains establish that, like $[PSI^+]$, $[MOT3^+]$ provides phenotypic diversity by altering the manifestation of natural genetic variation.

Wild strains harbour additional prions

How commonly do wild strains harbour prions that can create such heritable phenotypic diversity? Lacking means of detecting them by SDD-AGE, we used a phenotypic approach: we measured the growth of wild strains before and after GdHCl curing, across the same conditions used for $[PSI^+]$ and $[MOT3^+]$ (Supplementary Fig. 4). To ensure that any such phenotypes did not arise from *de novo* mutations, we compared four colonies of each wild strain with four cured derivatives (in total testing 5,520 isolates across 12 conditions).

Remarkably, over a third of the original wild strains (255) had phenotypes that differed in the same way between all four parental wild strains and all four cured derivatives. Moreover, nearly half of the growth properties conferred by these GdHCl-curable heritable elements were beneficial (Supplementary Table 1). The wild strain collection was biased towards wine isolates derived from natural fermentations. But it also contained many samples from beer, soil, fruit, infected human patients and commercial strains recently subject to man-made selective pressures to enhance properties for baking and brewing. Curable phenotypes, both beneficial and detrimental, occurred in yeasts from all of these niches. Even among the limited number of conditions tested here, prion curing had mixed phenotypic consequences in 15% of the strains. Thus, like $[PSI^+]$ and $[MOT3^+]$,

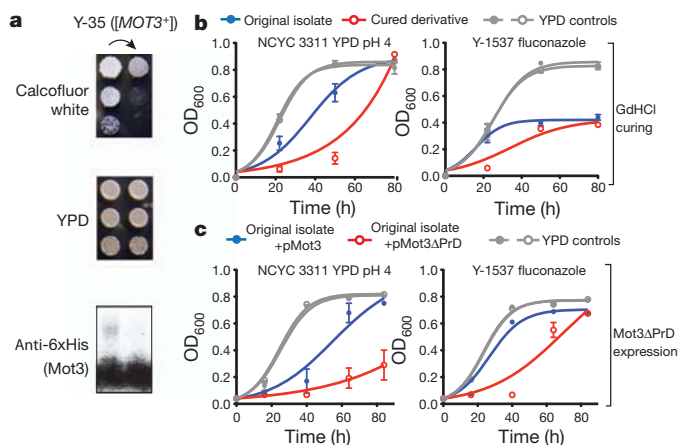


Figure 4 | Prions of the cell wall-remodelling transcription factor, Mot3, have diverse phenotypic consequences in wild strains. **a**, Strain Y-35 is resistant to the cell-wall-targeting drug, calcofluor white, but resistance is strongly reduced by passage on GdHCl. Probing for Mot3's endogenous hexahistidine motif reveals that Mot3 amyloids are retained in the uncured, but not the cured isolates. Apparent monomeric signal results from cross-reactivity to other yeast proteins. **b**, The $[MOT3^+]$ strains NCYC 3311 and Y-1537 (each shown in filled blue circles) are resistant to acidic growth conditions and fluconazole, respectively. Each phenotype is reversed by prion curing with GdHCl passage (open red circles). **c**, These phenotypes are also reversed by expression of a non-aggregating version of Mot3, Mot3 Δ PrD (open red circles). Expression of Mot3 itself (closed blue circles) did not affect the phenotype of either strain. Error bars represent the standard deviation of four independent biological replicates.

these prions created different trade-offs — traits that were beneficial or detrimental depending on circumstances.

To test whether the altered phenotypes arose from prion-mediated protein templating or from some unknown (yet somehow heritable and reproducible) effect of transient GdHCl exposure, we investigated 25 randomly chosen strains more rigorously. Transient expression of Hsp104^{DN} phenocopied the effects of GdHCl curing in 22 of the 25 strains (Supplementary Table 1), establishing their dependence on this prion-partitioning factor.

Another signature of prions is transmission to other cells via cytoplasmic transfer (cytoduction). Do the curable phenotypic elements of wild yeasts share this property? Because of the complexities in working with wild yeast, we used a derivative of W303, a common laboratory strain, as a universal 'recipient' for cytoplasmic transfer. Because prion-dependent traits can differ with genetic background, we chose wild strains with multiple curable traits as donors for these unknown prions, to increase the likelihood that transfer could be scored phenotypically. The South African wine strain WE372 had two traits that were heritably lost by prion curing (Fig. 5a, arrows): unusually robust growth on rich medium and poor growth at pH 9. The clinical isolate YJM428 had three such traits (Fig. 5a, arrows): robust growth in sodium chloride and 4-NQO, but poor growth on maltose.

After crossing donor and recipient strains to produce heterokaryons, we selected buds that bore only the nucleus of the W303 recipient but had received cytoplasm from the wild donor (Fig. 5b top; see Methods for details). We tested 12 such cytoductants from each mating to determine if they had inherited stable new traits from the cytoplasmic transfer.

Poor growth at pH 9 was not transferred, but robust growth on rich medium was transferred from WE372 donors to all 12 W303 recipients (Fig. 5c, red arrows). NaCl resistance was not transferred, but both enhanced growth on 4-NQO and poor growth on YP-maltose was transferred from YJM428 donors to all 12 W303 recipients (Fig. 5c, red arrows). The fidelity of the transferred traits established that they were not due to rare chromosome transfers that can occur in such crosses. The lack of transfer for some traits suggests that, as for $[PSI^+]$ and $[MOT3^+]$, traits produced by these unknown heritable cytoplasmic factors depend upon the genetic background. All transferred traits were curable by passage on GdHCl (Fig. 5c, black arrows), strongly indicating they were due to prions. Excluding the possibility that these traits were due to mitochondrial DNA transfer, we repeated the entire experiment with WE372 and YJM428 donors that had been cured of prions before heterokaryon formation (Fig. 5b, bottom). None acquired the new phenotypes (Fig. 5d).

Prions alter the relationship between genotype and phenotype in wild strains

How significantly do the prions of wild yeasts alter the phenotypic manifestation of genetic diversity? We examined the relationship between genotype and phenotype in the 21 strains in our collection whose genomes had been fully sequenced. As previously reported^{25,26}, the Spearman's correlation (ρ) between genotypes and phenotypes is typically on the order of 0.3 to 0.4 for wild yeast (in our strains and conditions, $\rho = 0.39$; $P = 3.5 \times 10^{-15}$). Prion loss made the correlation between genotype and phenotype weaker (Spearman's $\rho = 0.27$; $P = 1.5 \times 10^{-7}$). This finding was robust to random permutations of the data ($P = 0.0001$) and was clear even when $[PSI^+]$ and $[MOT3^+]$ strains were removed from the analysis. Thus, the prions these wild strains harbour broadly interface with polymorphisms in their genomes to influence the relationship between genotype and phenotype.

Discussion

The stable inheritance and complex phenotypes that prions produce arise from changes in protein conformation rather than from changes

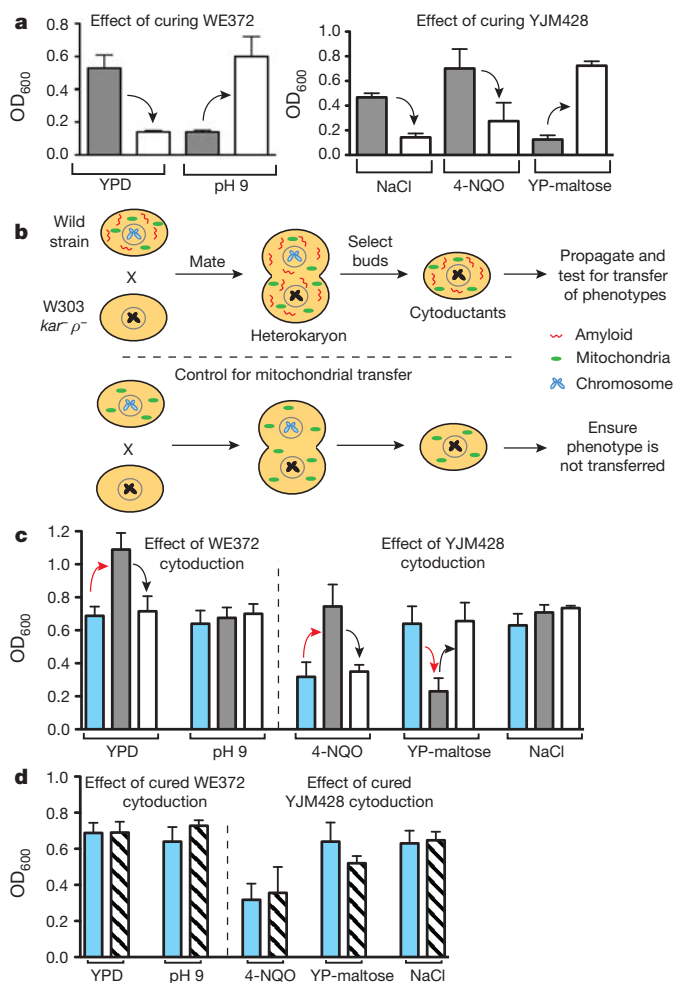


Figure 5 | The curable Hsp104-dependent epigenetic elements in wild yeast can be cytoplasmically transferred. **a**, Growth of wild strains WE372, YJM428, and their cured derivatives in selective conditions. Original isolates of each strain are shown in grey bars and their cured derivatives (indicated by black arrows) are shown in open bars. Error bars are one standard deviation from six biological replicates. **b**, Schematic for cytoduction experiments and control to ensure that phenotypes are due to prions, rather than transfer of mitochondrial DNA. **c**, Growth measurements of the laboratory recipient (blue bars), cytoductants (grey bars; red arrows denote cytoplasmic transfer of phenotypes), and cured derivatives of those cytoductants (open bars; black arrows denote curability of phenotypes). Error bars are the standard deviation of growth measurements from 12 cytoductants, the 12 cured derivatives of those cytoductants, or 6 biological replicates of the recipient strain. **d**, Cytoductants that received cytoplasm from cured derivatives of the original wild isolates (hashed bars) did not show an equivalent change in phenotype. Error bars represent the standard deviation of growth measurements from 12 cytoductants or 6 biological replicates of the recipient strain.

in nucleic acids. This non-canonical mode of inheritance has sparked considerable excitement and provoked intense mechanistic study (for review see ref. 27). But doubts about whether $[PSI^+]$ and other prions exist in nature have fuelled deep controversy over their relevance^{11,28}. We find that prions and prion-dependent phenotypes are widespread in nature, establishing their biological importance.

Saccharomyces cerevisiae is perhaps the most thoroughly characterized organism in experimental science. How, then, could this pervasive influence on the inheritance of biological traits have been missed for so long? The frequency of $[PSI^+]$ in wild strains suggests that previous efforts to find it simply did not examine enough strains (see Supplementary Information for further discussion). But we suspect that standard practices in yeast genetics provide a far more general explanation. Phenotypic analysis of new traits typically begins

by testing for 2:2 segregation in crosses and discarding variants that do not follow this Mendelian pattern. It is equally common to discard variants that prove to be restricted to individual strains. Thus, prion-based phenotypes may largely have been ignored because investigations were strongly biased by the known rules of nucleic-acid based inheritance and because of a pragmatism that neglected the biological significance of strain-to-strain variation.

We find that in wild strains a prion-dependent trait can readily be fixed by the meiotic re-assortment of endogenous genetic variation. (They can also probably be fixed by new mutations, a phenomenon not yet explored.) Thus, prions provide a robust mechanism by which yeast can increase their phenotypic diversity epigenetically, in a manner that readily allows that diversity to become hard-wired in subsequent generations. Evidence for an uncannily similar transition — from an epigenetically inherited trait to a genetically hard-wired trait — has recently surfaced for drug resistance in cancer cells^{29,30}. The underlying epigenetic mechanism in that case is chromatin-based rather than protein-based. But together, these observations point to a new view of the importance of epigenetics in evolutionary processes.

Under stressful conditions, cells increase the rate at which they switch into and out of the $[PSI^+]$ state¹⁰. This link between heritable phenotypic diversity and environmental contingency is a natural consequence of stress-induced disruptions in protein homeostasis^{31–34}. Such stress-regulated factors interface with protein homeostasis to drive increased prion switching in several ways (see Supplementary Information for discussion). Beyond these more general homeostatic mechanisms, it seems likely that conformational switches in individual prion proteins will prove to be regulated by additional protein-specific interactions and post-translational protein modifications.

Surprisingly, 40% of the traits produced by the wild prions we observed were beneficial to growth under the 12 conditions we tested. This contrasts with the consequences of mutational variants, the vast majority of which are either silent or detrimental. It seems probable, therefore, that the underlying, cryptic variation that creates prion-based traits in wild yeast, as well as the prions themselves, have been subject to previous selective events³⁵. In any case, the gain and loss of prions seems to constitute a sophisticated bet-hedging mechanism that allows cells to explore heritable new phenotypes more frequently in circumstances where they are not well suited to their environments.

Most of the 25 prions discovered to date are RNA-binding proteins, DNA-binding proteins and signal transducers — proteins that play key roles in governing the flow of information in cellular networks. Prion-mediated alterations in such functions allow access to complex traits in a single step. Together, the interface between prions, environmental stress, cryptic genetic variation and fixation provide a means to transition from a Lamarckian³⁶ mode of inheritance to a Darwinian framework of mutation and natural selection. As for Hsp90 (ref. 26), prions provide a robust route to the inheritance of environmentally induced traits that has moved from being merely plausible to demonstrable.

METHODS SUMMARY

SDD-AGE. Cells were grown for 18–24 h in 1 ml YPD followed by lysis in 96-well blocks. Lysates were treated with 2% SDS at room temperature and then clarified by centrifugation. Electrophoresis and blotting were performed as described³⁷.

Molecular cloning and yeast techniques. Standard cloning procedures¹⁴ were used for the construction of yeast plasmids marked with drug resistance cassettes. Yeast strains were obtained from the sources indicated in the Supplementary Information. Yeast handling, propagation and prion curing were as described¹⁴. Cytoductions used donor strains created by disrupting the *HO* locus with an antibiotic-resistance marker³⁸ and then sporulating. The recipient strain was a respiration deficient *KAR1-15* derivative of W303. High throughput phenotyping and data analyses were performed as described²⁶.

Sequence analysis of wild strains. Strains UCD978 and 5672 were sequenced to 100-fold coverage using the Illumina HiSeq platform. After quality control filtering, each genome was aligned against the *S. cerevisiae* reference sequence using

the BWA aligner³⁹. Correlation between similarity in genotype and similarity in phenotype was calculated as in ref. 26.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 12 November 2011; accepted 20 January 2012.

- True, H. L. & Lindquist, S. L. A yeast prion provides a mechanism for genetic variation and phenotypic diversity. *Nature* **407**, 477–483 (2000).
- Patino, M. M., Liu, J. J., Glover, J. R. & Lindquist, S. Support for the prion hypothesis for inheritance of a phenotypic trait in yeast. *Science* **273**, 622–626 (1996).
- Chernoff, Y. O., Lindquist, S. L., Ono, B., Inge-Vechtomov, S. G. & Liebman, S. W. Role of the chaperone protein Hsp104 in propagation of the yeast prion-like factor [psi⁺]. *Science* **268**, 880–884 (1995).
- Kryndushkin, D. S., Alexandrov, I. M., Ter-Avanesyan, M. D. & Kushnirov, V. V. Yeast [PSI⁺] prion aggregates are formed by small Sup35 polymers fragmented by Hsp104. *J. Biol. Chem.* **278**, 49636–49643 (2003).
- Chernoff, Y. O., Newnam, G. P., Kumar, J., Allen, K. & Zink, A. D. Evidence for a protein mutator in yeast: role of the Hsp70-related chaperone Ssb in formation, stability, and toxicity of the [PSI⁺] prion. *Mol. Cell. Biol.* **19**, 8103–8112 (1999).
- Lancaster, A. K., Bardill, J. P., True, H. L. & Masel, J. The spontaneous appearance rate of the yeast prion [PSI⁺] and its implications for the evolution of the evolvability properties of the [PSI⁺] system. *Genetics* **184**, 393–400 (2010).
- True, H. L., Berlin, I. & Lindquist, S. L. Epigenetic regulation of translation reveals hidden genetic variation to produce complex traits. *Nature* **431**, 184–187 (2004).
- Griswold, C. K. & Masel, J. Complex adaptations can drive the evolution of the capacitor [PSI⁺], even with realistic rates of yeast sex. *PLoS Genet.* **5**, e1000517 (2009).
- Lancaster, A. K. & Masel, J. The evolution of reversible switches in the presence of irreversible mimics. *Evolution* **63**, 2350–2362 (2009).
- Tyedmers, J., Madariaga, M. L. & Lindquist, S. Prion switching in response to environmental stress. *PLoS Biol.* **6**, e294 (2008).
- Halfmann, R., Alberti, S. & Lindquist, S. Prions, protein homeostasis, and phenotypic diversity. *Trends Cell Biol.* **20**, 125–133 (2010).
- Morimoto, R. I. Proteotoxic stress and inducible chaperone networks in neurodegenerative disease and aging. *Genes Dev.* **22**, 1427–1438 (2008).
- Masel, J. & Bergman, A. The evolution of the evolvability properties of the yeast prion [PSI⁺]. *Evolution* **57**, 1498–1512 (2003).
- Alberti, S., Halfmann, R., King, O., Kapla, A. & Lindquist, S. A systematic survey identifies prions and illuminates sequence features of prionogenic proteins. *Cell* **137**, 146–158 (2009).
- Crow, E. T. & Li, L. Newly identified prions in budding yeast, and their possible functions. *22*, 452–459 *Semin. Cell Dev. Biol.*, (2011).
- Halfmann, R. & Lindquist, S. Epigenetics in the extreme: prions and the inheritance of environmentally acquired traits. *Science* **330**, 629–632 (2010).
- McGlinchey, R. P., Kryndushkin, D. & Wickner, R. B. Suicidal [PSI⁺] is a lethal yeast prion. *Proc. Natl Acad. Sci. USA* **108**, 5337–5341 (2011).
- Nakayashiki, T., Kurtzman, C. P., Edskes, H. K. & Wickner, R. B. Yeast prions [URE3] and [PSI⁺] are diseases. *Proc. Natl Acad. Sci. USA* **102**, 10575–10580 (2005).
- Wickner, R. B. *et al.* Prion diseases of yeast: amyloid structure and biology. *22*, 469–475 *Semin. Cell Dev. Biol.*, (2011).
- Giudici, P. & Kunkee, R. E. The effect of nitrogen deficiency and sulfur containing amino acids on the reduction of sulfate to hydrogen sulfide by wine yeasts. *Am. J. Enol. Vitic.* **45**, 107–112 (1994).
- Derkatch, I. L. *et al.* Dependence and independence of [PSI⁺] and [PIN⁺]: a two-prion system in yeast? *EMBO J.* **19**, 1942–1952 (2000).
- Sondheimer, N. & Lindquist, S. Rnq1: an epigenetic modifier of protein function in yeast. *Mol. Cell* **5**, 163–172 (2000).
- Resende, C. G., Outeiro, T. F., Sands, L., Lindquist, S. & Tuite, M. F. Prion protein gene polymorphisms in *Saccharomyces cerevisiae*. *Mol. Microbiol.* **49**, 1005–1017 (2003).
- Chernoff, Y. O. *et al.* Evolutionary conservation of prion-forming abilities of the yeast Sup35 protein. *Mol. Microbiol.* **35**, 865–876 (2000).
- Liti, G. *et al.* Population genomics of domestic and wild yeasts. *Nature* **458**, 337–341 (2009).
- Jarosz, D. F. & Lindquist, S. Hsp90 and environmental stress transform the adaptive value of natural genetic variation. *Science* **330**, 1820–1824 (2010).
- Chien, P., Weissman, J. S. & DePace, A. H. Emerging principles of conformation-based prion inheritance. *Annu. Rev. Biochem.* **73**, 617–656 (2004).
- Wickner, R. B., Edskes, H. K., Shewmaker, F. & Nakayashiki, T. Prions of fungi: inherited structures and biological roles. *Nature Rev. Microbiol.* **5**, 611–618 (2007).
- Sharma, S. V. *et al.* A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. *Cell* **141**, 69–80 (2010).
- Feinberg, A. P. & Irizarry, R. A. Evolution in health and medicine Sackler colloquium: Stochastic epigenetic variation as a driving force of development, evolutionary adaptation, and disease. *Proc. Natl Acad. Sci. USA* **107** (suppl. 1), 1757–1764 (2010).
- Song, Y. *et al.* Role for Hsp70 chaperone in *Saccharomyces cerevisiae* prion seed replication. *Eukaryot. Cell* **4**, 289–297 (2005).
- Newnam, G. P., Wegryn, R. D., Lindquist, S. L. & Chernoff, Y. O. Antagonistic interactions between yeast chaperones Hsp104 and Hsp70 in prion curing. *Mol. Cell. Biol.* **19**, 1325–1333 (1999).
- Chernoff, Y. O. Stress and prions: lessons from the yeast model. *FEBS Lett.* **581**, 3695–3701 (2007).
- Chernova, T. A. *et al.* Prion induction by the short-lived, stress-induced protein Lsb2 is regulated by ubiquitination and association with the actin cytoskeleton. *Mol. Cell* **43**, 242–252 (2011).
- Masel, J. Cryptic genetic variation is enriched for potential adaptations. *Genetics* **172**, 1985–1991 (2006).
- Koonin, E. V. & Wolf, Y. I. Is evolution Darwinian or/and Lamarckian? *Biol. Direct* **4**, 42 (2009).
- Halfmann, R. & Lindquist, S. Screening for amyloid aggregation by semi-denaturing detergent-agarose gel electrophoresis. *J. Visualized Exp.* **17**, e838 (2008).
- Schacherer, J., Shapiro, J. A., Ruderfer, D. M. & Kruglyak, L. Comprehensive polymorphism survey elucidates population structure of *Saccharomyces cerevisiae*. *Nature* **458**, 342–345 (2009).
- Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to L. Bisson and L. Joseph, who provided the Department of Viticulture and Enology yeast collection from the University of California, Davis, USA. We also received strains from G. Fink, E. Louis, F. Dietrich and S. Dietzmann, and L. Kruglyak. We thank M. Taipale, K. Allendoerfer, K. Matlack, L. Pepper, V. Khurana, G. Fink, L. Joseph, A. Hochwagen and G. Walker for materials, discussions, and/or critical reading of the manuscript. N. Azubuine and T. Nanchung provided a constant supply of plates and media. S.L. is a Howard Hughes Medical Institute (HHMI) investigator. This work was supported by grants from the G. Harold and Leila Y. Mathers Foundation and HHMI. D.F.J. was supported as an HHMI fellow of the Damon Runyon Cancer Research Foundation (DRG-1964-08) and by an NIH Pathway to independence award (K99 GM098600). S.K.J. was supported as a summer research student by the Howard Hughes Medical Institute (HHMI) Exceptional Research Opportunities Program (EXROP).

Author Contributions R.H., D.F.J., A.K.L. and S.L. designed the experiments. R.H. performed SDD–AGE analyses. R.H., D.F.J., S.K.J. and A.C. carried out phenotyping experiments. D.F.J. analysed high-throughput phenotyping data. D.F.J. and A.K.L. analysed whole-genome sequence data. D.F.J. and A.C. performed cytoductions. R.H., D.F.J. and S.L. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.L. (Lindquist_admin@wi.mit.edu).

METHODS

Sequencing of wild strains. We sequenced two wild [*PSI*⁺] strains (UCD978 and 5672), a clinical isolate (YJM653), a wine strain (I14) and two lab isolates (W303 and YDJ25, a strain almost completely isogenic to the laboratory reference S288C strain). Using an Illumina HiSeq platform we obtained two lanes of 76 base pair paired-end reads and one lane of 101 base pair paired-end reads, resulting in an average coverage of 100-fold (all reads will be available at NCBI under accession number PRJNA81619). After quality control filtering, each genome was aligned against the *S. cerevisiae* reference sequence (sacCer2, June 2008 assembly, downloaded from UCSC on April 1, 2011: <http://hgdownload.cse.ucsc.edu/goldenPath/sacCer2/chromosomes/>) using the BWA aligner³⁹ followed by SNP and indel calling with respect to the reference using mpileup from the samtools⁴⁰ package.

We then estimated the genetic distance between each unique pair of genomes using a combination of custom code and the Genome Analysis Toolkit (GATK)⁴¹. For each genome pair we considered the superset of SNPs in both genomes: if a SNP was present only in one genome, but not the other, it was retained. SNPs that were present in both genomes were discarded. We constructed a neighbour-joining tree using the APE R package⁴² by using the genetic distance matrix from the counts of retained SNPs with quality score of at least 150 as called by mpileup. The horizontal scale bar in Supplementary Fig. 2 represents a genetic distance of 10,000 SNPs. To ensure that the tree was not sensitive to choice of the SNP quality cutoff, we performed the analysis at different quality thresholds and obtained essentially identical results.

We also separately sequenced the prion-determining region of *SUP35*. Only one had a non-synonymous change in the PrD, and this change was unlikely to influence Sup35's inherent prion propensity. All polymorphisms in the region spanning nucleotides -338 to 1102 of *SUP35* are indicated in Supplementary Table 2.

Molecular cloning. Standard cloning procedures were used¹⁴. Gateway destination vectors were constructed as follows. The *URA3* ORF in pAG416-GPD-*ccdB*, pAG416-TEF-*ccdB* and pAG416-*SUP35-ccdB*^{14,43} was replaced with cassettes conferring G418 or hygromycin B resistance from plasmids pUG6 and pAG32, respectively, to generate pAG41NEO-GPD-*ccdB*, pAG41NEO-*SUP35-ccdB* and pAG41HPH-TEF-*ccdB*. These plasmids contain the *GPDI*, *SUP35* and *TEF2* promoters, respectively, for driving the expression of exogenous genes. Gateway entry clones bearing the coding sequences for *SUP35*, *SUP35ΔPrD*, *HSP104*, *HSP104*^{KTKT} and *MOT3* were constructed as described¹⁴ using S288C genomic DNA and a plasmid bearing *HSP104* K218T K620T (ref. 3) as PCR templates. Site-directed mutagenesis was used to delete the PrD (amino acids 8–157) from the *MOT3* entry clone (R.H., personal communication). Entry clones and destination vectors were recombined in Gateway reactions to yield pAG41NEO-GPD-*HSP104*, pAG41NEO-GPD-*HSP104*^{KTKT}, pAG41NEO-*SUP35-SUP35*, pAG41NEO-*SUP35-SUP35ΔPrD*, pAG41HPH-TEF-*MOT3* and pAG41HPH-TEF-*MOT3ΔPrD*.

Yeast techniques. Yeast strains (Supplementary Table 1) were obtained from stock centres or generously provided by the sources indicated. All strains were stored as glycerol stocks at -80 °C and revived on YPD before testing. Strains that were [*PSI*⁺] in the original screen were re-ordered individually from the Department of Viticulture and Enology collection (University of California Davis) and the prion status was verified on a second SDD-AGE. Yeast were grown in YPD at 30 °C unless indicated otherwise. The following media supplements were included where relevant: 3 mM GdHCl, 200 μg ml⁻¹ G418, or 250 μg ml⁻¹ hygromycin B. Yeast were transformed with a standard lithium-acetate protocol as described⁴⁴.

To eliminate prions chemically, strains were passaged four times on rich medium containing 3 mM GdHCl. To eliminate prions by overexpression of *Hsp104*, cells were transformed with plasmids expressing *Hsp104* (wild type or K218T, K620T) from a strong constitutive promoter (GPD). Transformants were passaged three times on selective media, followed by four passages on YPD to allow plasmid loss, which was confirmed by the absence of growth on selective media. Finally, as GdHCl is known to increase the frequencies of petites, all GdHCl- and *Hsp104*^{DN}-cured isolates used in the phenotyping experiments for [*PSI*⁺], [*MOT3*⁺], and the analyses of 25 strains containing unknown prions, were first checked for respiration competence on glycerol. Cured and pre-cured isolates grew comparably well on glycerol.

A PCR-based deletion strategy⁴⁵ was used to replace one genomic copy of *URA3* in strain UCD978 with an *hphMX4* module from pAG32. The resulting strain was then sporulated. Random sporulants recovered on YPD containing hygromycin B were then sporulated again. Hygromycin-resistant sporulants from the second round of sporulation were tested for ploidy by growth on media containing 5-fluoroorotic acid (5-FOA), inability to grow on media lacking uracil, and ability to mate with haploid tester strains. Mating type was observed to be stable, indicating that UCD978 is heterothallic. Genomic DNA content was

determined by SYTOX Green staining as described⁴⁶, using a BD LSR II flow cytometer. BY4741 and BY4743 were used as haploid and diploid references, respectively.

For cytoduction experiments, we used a derivative of the common lab strain, W303, as a recipient for cytoplasmic transfer (Fig. 5b). The strain carried a dominant *KAR1-15* allele, which prevents nuclear fusion during mating but allows cytoplasmic transfer. The strain also carried multiple auxotrophic markers and a mitochondrial DNA defect. This allowed cytoplasmic transfer to be scored through the restoration of mitochondrial respiration, in the absence of transfer of auxotrophic markers. Haploid, mating-competent derivatives of the wild 'donor' strains were created using an antibiotic-resistance marker to knock out the *HO* locus³⁸. (*HO* would otherwise preclude mating by causing haploid cells to self-mate). The recipient and donor strains were patched together on rich media, followed by selection of heterokaryons on dropout media containing glycerol as a carbon source.

[RNQ⁺] strain UCD664 was verified to be *Saccharomyces uvarum* by colony PCR amplification and sequencing of the rDNA internal transcribed spacer region using oligonucleotides ITS1 and ITS4, as described⁴⁷. Another prion-containing strain not annotated as *S. cerevisiae*, UCD587, was originally annotated "*S. cerevisiae* race *bayanus*". We ordered this strain two independent times from the stock centre and found it to be *S. cerevisiae* based on *ITS1* sequence and growth characteristics.

Phenotypic assays. For agar adhesion and invasion analyses, colonies were allowed to grow for 5–7 days at 30 °C. Plates were then gently rinsed under running water to remove non-adherent cells; then photographed. The agar surface was then gently rubbed with a gloved finger under running water to remove all remaining non-invasive cells.

For growth measurements, wild yeast strains were inoculated in quadruplicate into 384-well plates containing 40 μl YPD per well and grown to saturation at 30 °C (typically 48 h) in a humidified chamber. After complete re-suspension, QRep polypropylene 384-well pin replicators (Genetix) were used to transfer cells (200–500 nl) to new 384-well plates in duplicate. These new plates contained 40 μl per well of selective media: with alternative carbon sources at 2% final concentration (YP-maltose, YP-galactose, YP-glycerol or YP-raffinose), compounds dissolved in YPD (0.5 M NaCl, 5% ethanol, 0.4 mg ml⁻¹ 4-NQO, 1 mM tBOOH, 32 mg ml⁻¹ fluconazole or 50 mM hydroxyurea), acidic and basic conditions (YPD at pH 4 or pH 9), and finally in YPD alone (as a control). Concentrated stocks of 4-NQO (4 mg ml⁻¹) were made in molecular biology grade dimethyl sulphoxide (DMSO) and stocks of fluconazole (64 mg ml⁻¹) were made in ethanol. All plates were incubated at 30 °C, covered and in a humidified chamber. Growth was measured approximately every 20 h by OD₆₀₀ with a Tecan Sapphire2 plate reader (after complete re-suspension by gentle agitation). Hits were chosen if all four replicates showed a significant change in growth (*P* < 0.01 determined by *t*-test) for at least two consecutive time points after curing of *Hsp104*-dependent prions.

Growth rates of [*PSI*⁺]-containing wild strains (and their cured derivatives) in conditions used for laboratory propagation (YPD, YM broth, FM broth, wort agar or Wallerstein nutrient agar) was measured in quadruplicate, diluting 10⁵ exponentially-growing cells of each strain in humidified 96-well plates containing 150 μl of medium. OD₆₀₀ of each well was measured every 15 min after re-suspension by agitation (15 s) in a plate reader (Multiskan Go, Thermo Scientific). Plates were incubated at 30 °C. After 4 days no appreciable loss in volume was observed in the exterior wells of the plates, ensuring that the growth rates we measured were not an artefact of evaporation. The exponential phase of plots of OD₆₀₀ versus time were fit to determine growth rates.

Correlation between similarity in genotype and similarity in phenotype was calculated in the R statistical computing software package using published genotypic correlation among sequenced strains and linear regression across the conditions in Supplementary Table 1. Significance was established by random perturbation of the phenotype data. Ten thousand permutations of the data did not reveal a similarly large change in genotype phenotype correlation occurring by chance.

SDD-AGE. Except where indicated below, SDD-AGE was performed as follows. Yeast were inoculated into 1 ml YPD in 96-well round-bottom blocks (Nunc P8241) with each well containing a single 3-mm glass bead. The blocks were incubated 18–24 h at 30 °C with 220 r.p.m. agitation. Cells were collected by centrifugation at 3,000 r.c.f. for 2 min, resuspended in 200 μl water, and then centrifuged again. Approximately 100 μl of acid-washed glass beads were then added to each well followed by 80 μl lysis buffer (100 mM Tris pH 8, 1% Triton X-100, 50 mM β-mercaptoethanol, 3% HALT protease inhibitor cocktail, 30 mM N-ethylmaleimide, and 12.5 U ml⁻¹ Benzonase nuclease). Blocks were then sealed with a rubber mat (Nunc 276002) and shaken at max speed twice for 3 min on a Qiagen Tissuelyser II. To each well was then added 35 μl 4× sample buffer (2× TAE, 20% glycerol, 8% SDS, 0.01% bromophenol blue). The blocks were then vortexed briefly and allowed to incubate at room temperature for three

minutes, followed by centrifugation for 2 min at 3,000 r.c.f. to remove cell debris. Electrophoresis and capillary blotting to Hybond ECL nitrocellulose were performed as described³⁷.

Sup35 and Rnq1 were detected using well-characterized antibodies¹⁴. Detection of amyloids using the Sup35 antibody was markedly improved by treating the blots with Re-Blot Plus stripping solution (Millipore) before probing. Available antibodies to yeast prion proteins Ure2, Swi1 and Cyc8 did not satisfactorily distinguish lab strains containing amyloids of these proteins from those that did not. Consequently, we did not attempt to identify these prions in wild strains. Horseradish-peroxidase-conjugated secondary antibodies were detected with Lumigen TMA-6 chemiluminescent substrate (GE).

For the detection of Mot3 amyloids, cells and lysates were prepared as described³⁷. Blots were probed with a hexahistidine antibody (GE Biosciences), which recognizes an endogenous hexahistidine motif in Mot3. From multiple experiments with a well characterized [MOT3⁺] isolate in the S288C genetic background (ref. 14 and R.H., personal communication), we estimated false-negative and false-positive rates with this antibody to be approximately 40% and 5%, respectively, under the conditions used here. We occasionally observed the appearance of [mot3⁻] colonies (as determined by SDD-AGE and phenotype) upon restreaking of [MOT3⁺] isolates. Polymorphisms in MOT3 (ref. 25) as

well as differences in the growth and spheroplasting efficiencies of wild strains may influence the validity of SDD-AGE for assessing Mot3's prion status. Nevertheless, phenotypic differences verified our designations of [MOT3⁺] for three of six strains identified by SDD-AGE.

40. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
41. DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genet.* **43**, 491–498 (2011).
42. Paradis, E., Claude, J. & Strimmer, K. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* **20**, 289–290 (2004).
43. Alberti, S., Gitler, A. D. & Lindquist, S. A suite of Gateway cloning vectors for high-throughput genetic analysis in *Saccharomyces cerevisiae*. *Yeast* **24**, 913–919 (2007).
44. Gietz, D., St Jean, A., Woods, R. A. & Schiestl, R. H. Improved method for high efficiency transformation of intact yeast cells. *Nucleic Acids Res.* **20**, 1425 (1992).
45. Goldstein, A. L. & McCusker, J. H. Three new dominant drug resistance cassettes for gene disruption in *Saccharomyces cerevisiae*. *Yeast* **15**, 1541–1553 (1999).
46. Haase, S. B. & Reed, S. I. Improved flow cytometric analysis of the budding yeast cell cycle. *Cell Cycle* **1**, 117–121 (2002).
47. Diezmann, S. & Dietrich, F. S. *Saccharomyces cerevisiae*: population divergence and resistance to oxidative stress in clinical, domesticated and wild isolates. *PLoS ONE* **4**, e5317 (2009).

Crystal structure of the channelrhodopsin light-gated cation channel

Hideaki E. Kato¹, Feng Zhang², Ofer Yizhar², Charu Ramakrishnan², Tomohiro Nishizawa¹, Kunio Hirata³, Jumpei Ito⁴, Yusuke Aita⁴, Tomoya Tsukazaki¹, Shigehiko Hayashi⁵, Peter Hegemann⁶, Andrés D. Maturana⁴, Ryuichiro Ishitani¹, Karl Deisseroth² & Osamu Nureki¹

Channelrhodopsins (ChRs) are light-gated cation channels derived from algae that have shown experimental utility in optogenetics; for example, neurons expressing ChRs can be optically controlled with high temporal precision within systems as complex as freely moving mammals. Although ChRs have been broadly applied to neuroscience research, little is known about the molecular mechanisms by which these unusual and powerful proteins operate. Here we present the crystal structure of a ChR (a C1C2 chimera between ChR1 and ChR2 from *Chlamydomonas reinhardtii*) at 2.3 Å resolution. The structure reveals the essential molecular architecture of ChRs, including the retinal-binding pocket and cation conduction pathway. This integration of structural and electrophysiological analyses provides insight into the molecular basis for the remarkable function of ChRs, and paves the way for the precise and principled design of ChR variants with novel properties.

Organisms ranging from archaeobacteria to human beings capture energy and/or information contained within environmental sources of light by using photoreceptors called rhodopsins, which consist of seven-transmembrane-helix proteins, called opsins, covalently linked to retinal. On the basis of primary sequences, the corresponding opsin genes are classified into two groups: microbial (type I) and animal (type II). Type I opsin genes are found in archaea, eubacteria, fungi and algae, whereas type II opsin genes are expressed in animals, including human beings. The type II proteins indirectly influence transmembrane ion currents by coupling to G-protein-based signal transduction pathways. In contrast, the type I proteins (not normally found in animals) include direct-light-activated regulators of transmembrane ion conductance, such as the light-driven ion pumps called bacteriorhodopsins and halorhodopsins (BRs and HRs)^{1,2} and the light-driven ion channels ChRs³. The light-driven ion pumps have been extensively studied, and structure–function relationships are well known. As opposed to these ion pumps, very little is known about the structure of ChRs or the mechanism by which these seven-transmembrane proteins conduct cations in a light-dependent manner³.

Beginning in 2005, it was found that ChRs could be expressed in mammalian neurons to mediate precise and reliable control of action potential firing in response to light pulses, without the need for exogenous retinal in vertebrate systems^{4–9}. ChRs have now been used to control neuronal activity in a wide range of animals, resulting in insights into fundamental aspects of circuit function as well as dysfunction and treatment in pathological states^{10,11}. However, despite the rapid progress of optogenetics (a technology also encompassing the use of ion pumps, such as HRs), virtually nothing is known about how a seven-transmembrane protein can form a light-switchable channel for cations. Although a rough helical arrangement was visible in the recently published ChR2 electron microscopy structure of two-dimensional crystals at 6 Å resolution, amino acid positioning and insights into channel function remained completely lacking¹². A high-resolution three-dimensional image would be of enormous value, not only to enhance understanding of microbial opsin-based channels,

but also to guide optogenetics in the generation of ChR variants with novel function related to spectrum, selectivity and kinetics. Even with limited structural models, ChR variants have been engineered with faster or extended open-state lifetimes^{13–17}, shifted absorption spectra^{13,17,18}, reduced desensitization^{6,18–20}, and increased expression and photocurrent magnitude^{6,15–17,19}. These advances represent the tip of the iceberg in terms of what could be achieved for all of the above properties, as well as for altered ion selectivity and unitary (single-channel) conductance properties, if detailed structural knowledge could be obtained to facilitate true electrostatic calculations and molecular dynamics simulations.

Here we present the ChR crystal structure at 2.3 Å resolution. This high-resolution information, along with electrophysiological analyses, has revealed the fundamentals of ChR architecture and guides the way to a basic working model for channelrhodopsin function.

Overall ChR structure

ChR2 from *C. reinhardtii* consists of 737 amino acids; the seven transmembrane domains (TMs) and photocurrent functionality are all contained within the amino-terminal ~300 amino acids. To identify the most promising candidates for structural studies, we constructed and explored an extensive range of different ChRs and ChR chimaeras with distinct carboxy-terminal truncations. Using fluorescence-detection size-exclusion chromatography (FSEC)²¹, we found that a novel chimaeric and truncated sequence termed C1C2, primarily consisting of ChR1 (ref. 3) without its C terminus and with the last two TMs swapped for those from ChR2 (related to previous chimaeras^{19,20,22} but with an additional six-amino-acid modification of the C terminus, namely removal of the sequence NKGTKG), was not only expressed well in Sf9 insect cells but also showed good stability and monodispersity as well as similar spectral characteristics to previous chimaeras²² (Supplementary Fig. 1). The crystals obtained from fully dark-adapted C1C2 in the lipidic cubic phase belonged to the C222₁ space group and diffracted X-rays to 2.3 Å resolution. We solved the C1C2 structure by the multiple anomalous dispersion (MAD)

¹Department of Biophysics and Biochemistry, Graduate School of Science, The University of Tokyo, 2-11-16 Yayoi, Bunkyo-ku, Tokyo 113-0032, Japan. ²Department of Bioengineering and Howard Hughes Medical Institute, Stanford University, Stanford, California 94305, USA. ³RIKEN SPring-8 Center, Hyogo 679-5148, Japan. ⁴Bioengineering Department, Nagaoka University of Technology, Niigata 940-2188, Japan. ⁵Department of Chemistry, Graduate School of Science, Kyoto University, Kyoto 606-8502, Japan. ⁶Institute of Biology, Experimental Biophysics, Humboldt-University, Invalidenstrasse 42, D-10115 Berlin, Germany.

method, using mercury-derivatized crystals (Supplementary Fig. 2). As far as we know, this is the first example of the phase determination by MAD for the crystal obtained in the lipidic cubic phase.

The truncated C1C2 chimera (residues 1–342) is composed of an N-terminal extracellular domain (N domain, residues 24–83, marked in Fig. 1a, d), the seven TMs (TM1–TM7; residues 84–317) connected by three cytoplasmic loops (ICL1–ICL3) and three extracellular loops (ECL1–ECL3), and the C-terminal intracellular domain (C domain, residues 318–356) (Fig. 1b, d). In addition to the region spanning the N-terminal residues 1–23, which is processed as a signal peptide (data not shown), residues 24–48, 110–117 and 343–356 are structurally disordered and invisible in the electron density map, whereas the core transmembrane region is clearly resolved (Fig. 1a). Searches of the Protein Data Bank using the Dali server (<http://ekhidna.biocenter.helsinki.fi/dali>) suggested that the N domain, consisting of a short 3_{10} -helix and two β -strands, has a novel fold. Within each C1C2 protomer, 6 lipids and 43 water molecules were observed.

Two C1C2 protomers were found to be tightly associated into a closely apposed dimer, as previously predicted from electron microscopy¹². Interfacial interactions occur in the N domain, ECL1, TM3 and TM4 of each molecule (Fig. 1b, c). Notably, Cys 66 (27), Cys 73 (34) and Cys 75 (36) in the N domain (ChR2 numbering is shown in parentheses here and below for comparison with earlier literature) form three disulphide bonds between protomers. As Cys 73 and Cys 75 are highly conserved in ChRs, this interaction may contribute to stabilizing the N-domain interaction and molecular dimerization (Supplementary Fig. 3).

Structural comparison with BR

We next compared the C1C2 structure with that of BR and bRh (bovine rhodopsin). The primary sequence of ChR is similar to that

of BR as well as other microbial opsins, such as xanthorhodopsin and sensory rhodopsin II (Supplementary Fig. 4). Consistent with this sequence similarity, C1C2 superimposed well on BR (PDB accession 1IW6)²³, but not on bRh (PDB accession 3C9L)²⁴ (Fig. 1d and Supplementary Fig. 5). TM3–6 between C1C2 and BR are very similar, and the position of the protonated Schiff base is conserved (Fig. 1e), whereas there are three distinct features between the two structures. First, C1C2 has additional N-terminal and C-terminal domains. The N domain, as noted above, contributes to dimer formation, and the C domain may be involved in subcellular localization and scaffolding in *Chlamydomonas*; for example, to tether ChR to the algal eyespot²⁵. Second, TM7 of C1C2 protrudes into the intracellular space, projecting ~ 18 Å from the membrane surface, and the intracellular end of TM7 is shifted towards the central axis of the monomer by 2.7 Å, as compared with BR (Fig. 1d). Although the function of this protruding part of TM7 is unclear, His 313, His 317 and Gly 318 may contribute to stabilizing the intracellular C domain via a water-mediated hydrogen-bonding network (Supplementary Fig. 6). Last, and most importantly, the C1C2 extracellular ends of TM1 and TM2 are tilted outward by 3.0 Å and 4.1 Å, respectively, compared to those of BR (Fig. 1e). These tilts enlarge the cavity formed by TM1, 2 and 7 and allow water influx for a cation-translocation pathway, as discussed later.

Retinal-binding pocket, Schiff base and counterion

As in other microbial-type rhodopsins, all-*trans* retinal (ATR) is covalently bound to Lys 296 (257) on TM7 (Lys 216 in BR), forming the Schiff base. As in BR, five aromatic residues (Trp 163, Phe 217, Trp 262, Phe 265 and Phe 269) are located around the polyene chain and the β -ionone ring, forming a hydrophobic pocket for ATR (Fig. 2a, b), whereas Cys 167 (128), Thr 198 (159) and Ser 295 (256) form a less-hydrophobic pocket, and may contribute to colour shift (Fig. 2a, b).

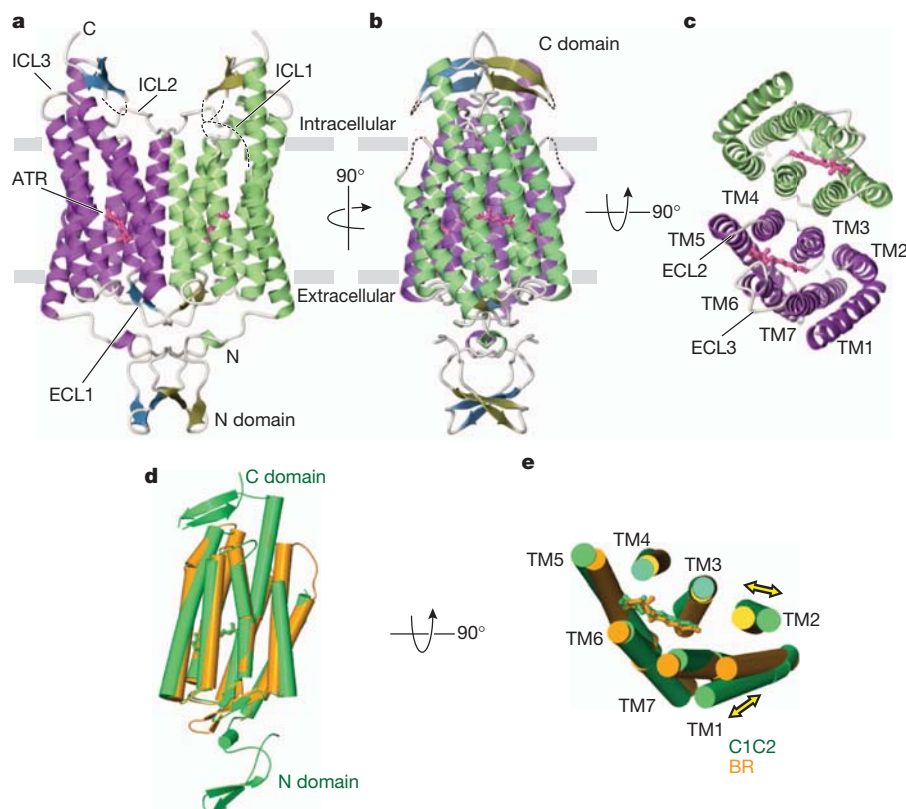


Figure 1 | Structure of C1C2 and comparison with BR. a–c, Crystal structure of the C1C2 dimer, viewed parallel to the membrane from two angles (a, b), and viewed from the extracellular side (c). C1C2 consists of the N domain, the seven transmembrane helices (TM1–TM7) connected by extracellular loops (ECL1–ECL3) and intracellular loops (ICL1–ICL3), and the C domain. Disordered

regions are represented as dotted lines. The ATR is coloured pink. d, e, Side view (d) and extracellular view (e) of the superimposed TMs of C1C2 (green) and BR (orange). The yellow double arrows indicate the shifts of the extracellular parts of TM1 and TM2.

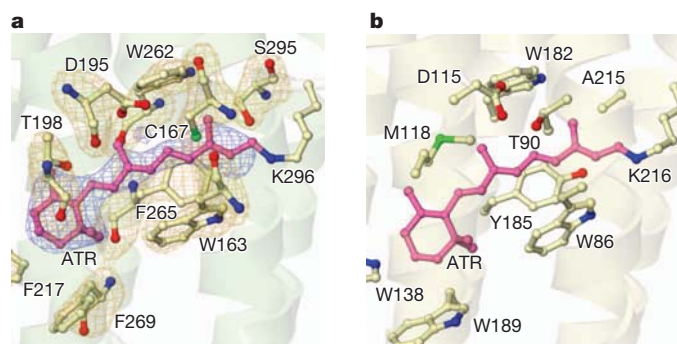


Figure 2 | Structural comparison of the retinal-binding pocket between C1C2 and BR. **a**, Structure of the retinal-binding pocket of C1C2, with an omit map of ATR at 3σ and of the surrounding residues (subtract 39 from the C1C2 residue number to obtain Chr2 numbering) at 3.5σ . **b**, Structure of the retinal-binding pocket of BR.

A previous report suggested that the side chains of Cys 167 (128) and Asp 195 (156) (Thr 90 and Asp 115 in BR) directly hydrogen bond with each other²⁶ (Fig. 2a), and that this interaction may function as the molecular switch to direct transition to the conducting state. However, in the present C1C2 structure, the distances between the thiol group of Cys 167 and the carboxyl oxygen atoms of Asp 195 are 4.4 Å and 4.6 Å, respectively, and the thiol group of Cys 167 is not associated with Asp 195, but with the π -electron system in the retinal molecule (Fig. 2a).

In BR, a water molecule receives a proton from the protonated Schiff base and donates a proton to Asp 85 (ref. 27; Fig. 3b); this arrangement is conserved in C1C2. However, in C1C2 the distances from the protonated Schiff base are 4.4 Å, 3.4 Å and 3.0 Å respectively for the water molecule, Glu 162 (123) (Asp 85 in BR) and Asp 292 (253) (Asp 212 in BR) (Fig. 3a, b). Therefore, in C1C2, Asp 292 or

possibly Glu 162, but not the water located between them, may directly receive a proton from the protonated Schiff base. In BR, Asp 212 retains a low pK_a because it is hydrogen bonded to Tyr 57 and Tyr 185, which do not move during the photocycle (Fig. 3b). On the other hand, in C1C2, Tyr 57 and Tyr 185 are substituted with Phe 133 and Phe 265, and Asp 292 only forms a hydrogen bond with water (Fig. 3a); thus Asp 292 could move relatively freely in the photocycle. Therefore in C1C2, the pK_a of Asp 292 can change, in contrast to the corresponding residue in BR. Moreover, the pK_a values of Glu 162 and Asp 292, calculated using PROPKA²⁸ (Supplementary Table 2), showed that Glu 162 may be protonated and Asp 292 may be deprotonated in our structure. Thus, we suggest that Asp 292, rather than Glu 162, is the primary proton acceptor in C1C2, consistent with the finding that Glu 123 mutants show current amplitudes similar to wild type^{13,17}. To verify further this notion, we expressed the E162A and D292A mutants of C1C2 in HEK293 cells, and recorded photocurrents in response to 465-nm light pulses (Fig. 3c, d and Supplementary Figs 7 and 8), revealing that replacement of Glu 162 by alanine resulted in moderately decreased currents, whereas the substitution of Asp 292 by alanine almost completely abolished photocurrents despite robust membrane expression. Moreover, the onset time constant (τ_{ON}) of the D292A mutant was significantly larger than that of wild type (Supplementary Fig. 9), consistent with the structure showing that Asp 292, rather than Glu 162, may be the major proton acceptor from the protonated Schiff base in ChR.

Electronegative pore framed by four TM helices

We calculated the electrostatic surface potential of C1C2, which revealed an electronegative pore formed by TM1, 2, 3 and 7 (Fig. 4a). In this pathway, a number of negatively charged residues, including Glu 129 (90), Glu 136 (97) and Glu 140 (101), as well as Glu 162 (123) and Asp 292 (253), are aligned along the pore (Fig. 4b). Because most of the negatively charged residues are derived

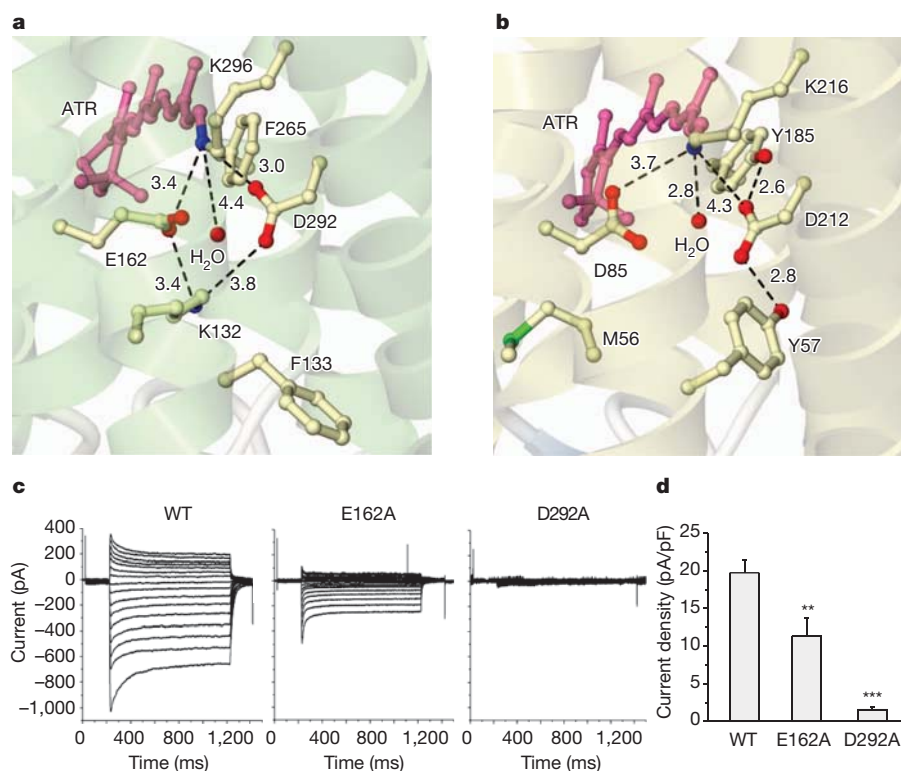


Figure 3 | The protonated Schiff base and its counterions in C1C2 and BR. **a**, **b**, Structures of the environment around the Schiff base in C1C2 (**a**) and BR (**b**). Numbers indicate the distance between two atoms connected by dashed lines. **c**, Effects of the mutation of two possible counterions on the

photocurrent. Photocurrents on C1C2-expressing HEK293 cells were measured at 16 different holding potentials. WT, wild type. **d**, The peak amplitudes of the photocurrents, normalized by the cell's input capacitance. Values are means and s.e.m. of 7–15 experiments. ** $P < 0.01$, *** $P < 0.001$.

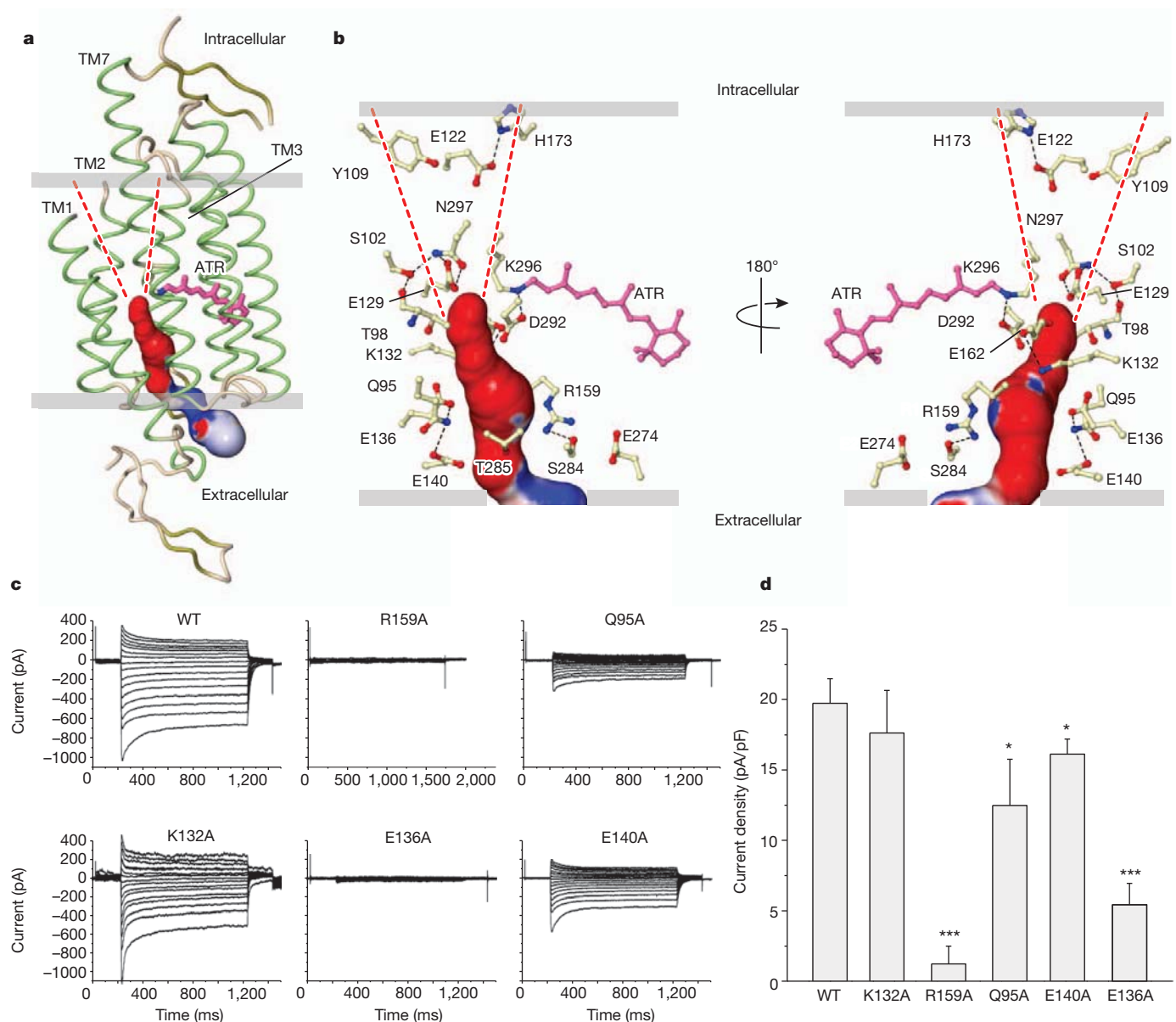


Figure 4 | Cation-conducting pathway formed by TM1, 2, 3 and 7. **a**, Pore-lining surface calculated by the CAVER³⁷ program, coloured by the electrostatic potential. Dashed red lines indicate putative intracellular vestibules. **b**, Close-up views of the surface of the pore, with the 17 polar lining residues (subtract 39 from the C1C2 residue number to obtain Chr2 numbering). Hydrogen bonds

are shown as black dashed lines. **c**, Photocurrents of mutants of the five residues within the pathway, measured under the same conditions as in Fig. 3c. **d**, The peak amplitudes of the photocurrents, as in Fig. 3d. * $P < 0.05$, *** $P < 0.001$. Error bars represent s.e.m.

from TM2, we suggest that the ion conductance and selectivity of C1C2 are mainly defined by TM2.

On the extracellular side of the pore, a vestibule formed by the N domain and ECL1–3 opens up to a diameter of about 8 Å, where Lys 154 (115), Lys 209 (170) and Arg 213 (174) form a slightly electropositive surface around the vestibule (Supplementary Fig. 10). Deeper within the vestibule, Arg 159 (120), Tyr 160 (121), Glu 274 (235) and Ser 284 (245) form a weak electronegative surface and fix the positions of TM3, 6 and 7 by a water-mediated hydrogen-bond network (Supplementary Fig. 11a). As these four residues are highly conserved not only in ChRs but also in BRs (Arg 82, Tyr 83, Glu 194 and Glu 204, respectively), and the corresponding residues in BR reportedly have an important role in proton pumping, we generated the R159A mutant in C1C2. We found that this mutant did not produce a photocurrent despite robust membrane expression (Fig. 4c, d and Supplementary Figs 7 and 8); because the orientation of Arg 159 is quite different from the corresponding Arg residue in

BR, and these residues form the extracellular hydrophilic surface, we suggest that this conserved cluster has an important role in creation of the extracellular vestibule, rather than in proton movement as in BR (Supplementary Fig. 11b).

In the middle of the pore, 12 polar residues (Gln 95 (56), Thr 98 (59), Ser 102 (63), Glu 122 (83), Glu 129 (90), Lys 132 (93), Glu 136 (97), Glu 140 (101), Glu 162 (123), Thr 285 (246), Asp 292 (253) and Asn 297 (258)) form a hydrophilic and strongly electronegative surface (Fig. 4b). To investigate contributions to ChR function, we measured photocurrents, kinetics and selectivity for four mutants (Q95A, K132A, E136A, E140A) (Fig. 4c, d and Supplementary Figs 7–9 and 12). The K132A mutant had faster kinetics and similar current amplitude relative to wild type, whereas the Q95A and E140A mutants exhibited moderately reduced currents, and the E136A mutant showed very little photocurrent, despite robust membrane expression. Three of the four mutants (Q95A, K132A and E136A) altered ion selectivity (Supplementary Fig. 12); therefore, we suggest that this

pore is important for cation conduction and that, as previously suggested²⁹, Glu 136 (97) is essential.

Although this putative cation-conducting pathway is opened towards the extracellular side, the cytoplasmic side of this pathway is occluded owing to two constrictions (Figs 4b and 5). The first constriction is formed by three highly conserved polar residues: Ser 102 (63), Glu 129 (90) and Asn 297 (258) (Figs 4b and 5a). In this constriction site, Ser 102, with a β -OH group that hydrogen bonds to the main-chain carbonyl oxygen of Thr 98, fixes the position of Asn 297 and, in turn, Asn 297 fixes Glu 129 by hydrogen bonds. Glu 129 protrudes into and occludes the pore. To analyse this putative channel gate, we prepared four mutants (S102D, E129A, E129Q and N297D) and measured photocurrents, kinetics and ion selectivity (Supplementary Figs 7–9, 12 and 13). We found that E129A, E129Q and N297D affect ion selectivity and S102D, E129Q and N297D affect channel kinetics (Supplementary Figs 9 and 12). These results, consistent with previous work^{30,31}, indicate the importance of these three residues and suggest that cations pass through this constriction site in the conducting state.

The second constriction is formed by the phenol group of Tyr 109 (70) (Figs 4b and 5b). Given the high *B*-factor of the C-terminal end of TM1 (Supplementary Fig. 14) and a previous Fourier transform infrared spectroscopy (FT-IR) study showing that the α -helices undergo conformational changes during the photocycle³², movement of the TM1 C-terminal end may open the pore exit formed between TM1, 2 and 7. As TM1 does not directly interact with the retinal chromophore, the signal of retinal isomerization is expected to be transmitted to TM1 via movements of TM2, 3 and/or 7. However, we cannot exclude the possibility that movements of TM2, 3 and/or 7 form a cytoplasmic vestibule next to Tyr 109, and further studies will be required to identify the pore exit.

Discussion

This first crystal structure of a light-gated cation channel in the closed/dark state at 2.3 Å resolution provides insight into ChR dimerization, retinal binding and cation conductance. Moreover, owing to the large N domain unique to ChR, it has been difficult to align precisely the ChR sequence with other microbial rhodopsins (notably BR), and the present structure-based alignment (Supplementary Fig. 4) will assist in the design and interpretation of functional analyses, including electrophysiological and spectroscopic studies of ChR at the molecular level.

The structural features around the ATR, Schiff base, and conduction pathway also provide insight into the blueshifted absorption spectrum of ChR ($\lambda_{\text{max}} = 470$ nm), as compared to that of BR ($\lambda_{\text{max}} = 568$ nm). In general, the maximum absorption wavelength of retinal proteins is

determined by the energy difference between ground (S_0) and excited (S_1) states, and this gap is mainly affected by the planarity of the conjugated system of the retinal chromophore, the distance between the protonated Schiff base and its counterion, and the interaction of the chromophore with polar or polarizable residues³³. Although the planarity of the ATR is unchanged between ChR and BR, the counterion of ChR, Asp 292, is located ~ 1 Å closer to the Schiff base than the corresponding Asp residue of BR (Fig. 3a, b), and the negatively charged residues are aligned along the conducting pathway (Fig. 4a, b). These environments are likely to stabilize the S_0 state of ChR, thus enlarging the energy gap between the S_0 and S_1 states and thereby causing the relative absorption spectrum blueshift.

Much about the photocycle remains unknown but is thought to be similar to that of BR^{31,34,35}, in which the essential early event is the dipole change of the protonated Schiff base, which alters the nitrogen pK_a by several orders of magnitude. In the case of ChRs, this may cause the release of the Schiff base proton to Asp 292 (probably not to either Glu 162 or water because Asp 292 is closer than these other two moieties, which also may explain why the D292A mutant is inactive; Fig. 3). The protonation of Asp 292 is likely to repel Lys 132 (93), as with Arg 82 in BR, and this movement may enlarge pore diameter and help cations to pass (Figs 3a and 4b). It is also thought that channel opening may be coupled with reprotonation of the Schiff base. Given the calculated pK_a (Supplementary Table 2) and the distance from the Schiff base nitrogen atom, we suggest two candidates for this proton donor—Glu 122 and Glu 129 (Supplementary Fig. 15)—but further studies, including structural determinations of photocycle intermediate states, will be required to refine our model.

In recent years, many strategies have been applied to create ChR variants with improved properties for optogenetics, ranging from designer ChR variants based on functional and structural similarities between BR and ChR (E123X, H134R, C128X, I131V, D156A, T159C, C1V1)^{6,13–15,17,19,36}, to chimaera construction along with mutagenesis (ChRGR, L132C, ChD, ChEF, C1V1)^{16–19} (Fig. 6a). These approaches have generated a number of ChR variants with useful properties, but the high-resolution crystal structure is a prerequisite for the design of ChR variants with ideal properties. The present crystal structure describes the environment around the retinal-binding pocket (Fig. 6a, b), which will enable optimized design of red- and blueshifted ChR variants. In addition, structure of the cation-conducting pathway may facilitate construction of ChR variants with improved photocurrents, photo-sensitivity, cation selectivity and kinetics. For example, K132A and Q95A show strong photocurrents and K^+ selectivity (Supplementary Figs 9 and 12), which could be useful to suppress neural activity.

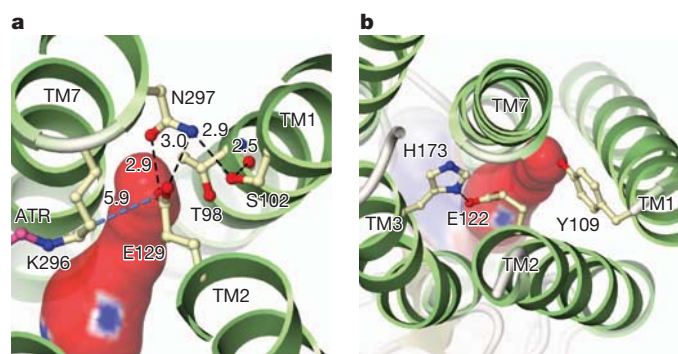


Figure 5 | Two constriction sites on the cytoplasmic side of C1C2 in the closed state. **a**, The first constriction site is formed by Ser 102 (63), Glu 129 (90) and Asn 297 (258). Hydrogen bonds are shown as black dashed lines. The blue dashed line represents a possible proton transfer pathway. **b**, The second constriction, made by Tyr 109. The cavity formed by TM1, 2 and 7 is occluded by Tyr 109, and the cavity formed by TM2, 3 and 7 is occluded by Glu 122 and His 173.

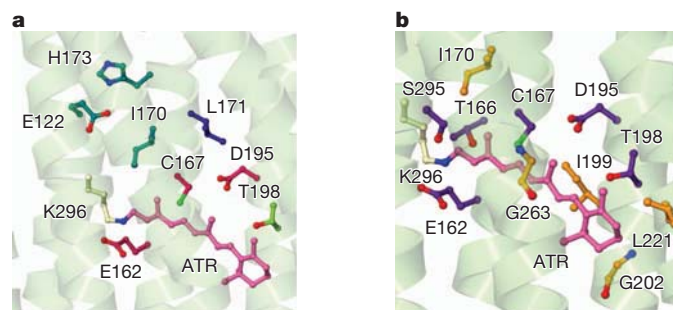


Figure 6 | Distribution of known mutations and possible candidates for future mutations. **a**, Mutations that affect both the absorption spectrum and the kinetics (Cys 167 (128), Glu 162 (123) and Asp 195 (156); deep red), the conductance (Thr 198 (159); light green), the selectivity (Leu 171 (132); dark blue) and the kinetics (Glu 122 (83), Ile 170 (131) and His 173 (134); dark cyan) of ChR2. **b**, Polar (Glu 162 (123), Thr 166 (127), Cys 167 (128), Asp 195 (156), Thr 198 (159) and Ser 295 (256); magenta) and non-polar (Ile 170 (131), Ile 199 (160), Gly 202 (163), Leu 221 (182) and Gly 263 (224); orange) residues surrounding ATR.

To understand the photocycle in more detail, further structural studies, including determination of crystal structures in intermediate states, are clearly needed. However, the present structural information represents a key step in enabling the principled design of ChR variants with new properties, and will accelerate both applications of optogenetics to intact-systems biology, and basic mechanistic understanding of these remarkable photoreceptor proteins.

METHODS SUMMARY

C1C2 was cloned into cleavable enhanced green fluorescent protein (EGFP)-His₆ fusion pFastBac1 vector. The fusion protein was expressed in insect cells, solubilized in 2.5% (w/v) *n*-dodecyl- β -D-maltoside (DDM) and 0.5% (w/v) cholesteryl hemisuccinate, and purified by nickel affinity chromatography. After that, the C-terminal EGFP was cleaved by His-tagged tobacco etch virus (TEV) protease (homemade). Then the sample mixture was passed through Ni-NTA resin again to remove the cleaved His-tagged EGFP and TEV protease. The sample was further purified by size-exclusion chromatography. Crystals were grown in a lipidic cubic phase using monoolein. Diffraction data were measured at beamline X06SA of the Swiss Light Source and at beamline BL32XU of SPring-8. The structure was solved by the MAD method using mercury derivative. Data collection and refinement statistics are presented in Supplementary Table 1. Electrophysiological recordings were conducted using patch-clamp on HEK293 cells expressing the wild-type and mutant C1C2.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 28 October 2011; accepted 3 January 2012.

Published online 22 January 2012.

- Oesterhelt, D. & Stoekenius, W. Rhodopsin-like protein from the purple membrane of *Halobacterium halobium*. *Nat. New Biol.* **233**, 149–152 (1971).
- Matsuno-Yagi, A. & Mukohata, Y. Two possible roles of bacteriorhodopsin; a comparative study of strains of *Halobacterium halobium* differing in pigmentation. *Biochem. Biophys. Res. Commun.* **78**, 237–243 (1977).
- Nagel, G. *et al.* Channelrhodopsin-1: a light-gated proton channel in green algae. *Science* **296**, 2395–2398 (2002).
- Boyden, E. S., Zhang, F., Bamberg, E., Nagel, G. & Deisseroth, K. Millisecond-timescale, genetically targeted optical control of neural activity. *Nature Neurosci.* **8**, 1263–1268 (2005).
- Li, X. *et al.* Fast noninvasive activation and inhibition of neural and network activity by vertebrate rhodopsin and green algae channelrhodopsin. *Proc. Natl Acad. Sci. USA* **102**, 17816–17821 (2005).
- Nagel, G. *et al.* Light activation of channelrhodopsin-2 in excitable cells of *Caenorhabditis elegans* triggers rapid behavioral responses. *Curr. Biol.* **15**, 2279–2284 (2005).
- Ishizuka, T., Kakuda, M., Araki, R. & Yawo, H. Kinetic evaluation of photosensitivity in genetically engineered neurons expressing green algae light-gated channels. *Neurosci. Res.* **54**, 85–94 (2006).
- Bi, A. *et al.* Ectopic expression of a microbial-type rhodopsin restores visual responses in mice with photoreceptor degeneration. *Neuron* **50**, 23–33 (2006).
- Zhang, F., Wang, L. P., Boyden, E. S. & Deisseroth, K. Channelrhodopsin-2 and optical control of excitable cells. *Nature Methods* **3**, 785–792 (2006).
- Yizhar, O., Fenno, L. E., Davidson, T. J., Mogri, M. & Deisseroth, K. Optogenetics in neural systems. *Neuron* **71**, 9–34 (2011).
- Fenno, L., Yizhar, O. & Deisseroth, K. The development and application of optogenetics. *Annu. Rev. Neurosci.* **34**, 389–412 (2011).
- Muller, M., Bamann, C., Bamberg, E. & Kuhlbrandt, W. Projection structure of channelrhodopsin-2 at 6 Å resolution by electron crystallography. *J. Mol. Biol.* **414**, 86–95 (2011).
- Gunaydin, L. A. *et al.* Ultrafast optogenetic control. *Nature Neurosci.* **13**, 387–392 (2010).
- Berndt, A., Yizhar, O., Gunaydin, L. A., Hegemann, P. & Deisseroth, K. Bi-stable neural state switches. *Nature Neurosci.* **12**, 229–234 (2009).
- Berndt, A. *et al.* High-efficiency channelrhodopsins for fast neuronal stimulation at low light levels. *Proc. Natl Acad. Sci. USA* **108**, 7595–7600 (2011).
- Kleinlogel, S. *et al.* Ultra light-sensitive and fast neuronal activation with the Ca²⁺-permeable channelrhodopsin CatCh. *Nature Neurosci.* **14**, 513–518 (2011).
- Yizhar, O. *et al.* Neocortical excitation/inhibition balance in information processing and social dysfunction. *Nature* **477**, 171–178 (2011).
- Wen, L. *et al.* Opto-current-clamp actuation of cortical neurons using a strategically designed channelrhodopsin. *PLoS ONE* **5**, e12893 (2010).
- Lin, J. Y., Lin, M. Z., Steinbach, P. & Tsien, R. Y. Characterization of engineered channelrhodopsin variants with improved properties and kinetics. *Biophys. J.* **96**, 1803–1814 (2009).
- Wang, H. *et al.* Molecular determinants differentiating photocurrent properties of two channelrhodopsins from *Chlamydomonas*. *J. Biol. Chem.* **284**, 5685–5696 (2009).
- Kawate, T. & Gouaux, E. Fluorescence-detection size-exclusion chromatography for precrystallization screening of integral membrane proteins. *Structure* **14**, 673–681 (2006).
- Tsunoda, S. P. & Hegemann, P. Glu 87 of channelrhodopsin-1 causes pH-dependent color tuning and fast photocurrent inactivation. *Photochem. Photobiol.* **85**, 564–569 (2009).
- Matsui, Y. *et al.* Specific damage induced by X-ray radiation and structural changes in the primary photoreaction of bacteriorhodopsin. *J. Mol. Biol.* **324**, 469–481 (2002).
- Stenkamp, R. E. Alternative models for two crystal structures of bovine rhodopsin. *Acta Crystallogr. D* **64**, 902–904 (2008).
- Mittelmeier, T. M., Boyd, J. S., Lamb, M. R. & Dieckmann, C. L. Asymmetric properties of the *Chlamydomonas reinhardtii* cytoskeleton direct rhodopsin photoreceptor localization. *J. Cell Biol.* **193**, 741–753 (2011).
- Nack, M. *et al.* The DC gate in Channelrhodopsin-2: crucial hydrogen bonding interaction between C128 and D156. *Photochem. Photobiol. Sci.* **9**, 194–198 (2010).
- Lanyi, J. K. Proton transfers in the bacteriorhodopsin photocycle. *Biochim. Biophys. Acta* **1757**, 1012–1018 (2006).
- Bas, D. C., Rogers, D. M. & Jensen, J. H. Very fast prediction and rationalization of pK_a values for protein-ligand complexes. *Proteins* **73**, 765–783 (2008).
- Sugiyama, Y. *et al.* Photocurrent attenuation by a single polar-to-nonpolar point mutation of channelrhodopsin-2. *Photochem. Photobiol. Sci.* **8**, 328–336 (2009).
- Ruffert, K. *et al.* Glutamate residue 90 in the predicted transmembrane domain 2 is crucial for cation flux through channelrhodopsin 2. *Biochem. Biophys. Res. Commun.* **410**, 737–743 (2011).
- Piazza, A. P. *et al.* Bioinformatic and mutational analysis of channelrhodopsin-2 cation conducting pathway. *J. Biol. Chem.* <http://dx.doi.org/10.1074/jbc.M111.326207> (2011).
- Radu, I. *et al.* Conformational changes of channelrhodopsin-2. *J. Am. Chem. Soc.* **131**, 7313–7319 (2009).
- Lasogga, L., Rettig, W., Otto, H., Wallat, I. & Bricks, J. Model systems for the investigation of the opsin shift in bacteriorhodopsin. *J. Phys. Chem. A* **114**, 2179–2188 (2010).
- Ritter, E., Stehfest, K., Berndt, A., Hegemann, P. & Bartl, F. J. Monitoring light-induced structural changes of Channelrhodopsin-2 by UV-visible and Fourier transform infrared spectroscopy. *J. Biol. Chem.* **283**, 35033–35041 (2008).
- Bamann, C., Kirsch, T., Nagel, G. & Bamberg, E. Spectral characteristics of the photocycle of channelrhodopsin-2 and its implication for channel function. *J. Mol. Biol.* **375**, 686–694 (2008).
- Bamann, C., Gueta, R., Kleinlogel, S., Nagel, G. & Bamberg, E. Structural guidance of the photocycle of channelrhodopsin-2 by an interhelical hydrogen bond. *Biochemistry* **49**, 267–278 (2010).
- Petrík, M. *et al.* CAVER: a new tool to explore routes from protein clefts, pockets and cavities. *BMC Bioinformatics* **7**, 316 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank Y. Tanaka, T. Higuchi, M. Hattori and H. Nishimasu for useful discussions; T. Hino for technical support; and the beamline staff members at BL32XU of SPring-8 (Hyogo, Japan) and at X06SA of the Swiss Light Source (Villigen, Switzerland) for technical help during data collection. This work was supported by the Japan Society for the Promotion of Science (JSPS) through its “Funding Program for World-Leading Innovative R&D on Science and Technology (FIRST program)” to O.N., by a grant for the National Project on Protein Structural and Functional Analyses from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) to O.N., and by a Grant-in-Aid for Scientific Research (S) from MEXT to O.N. F.Z. is supported by the McKnight Foundation. K.D. is supported by the Gatsby Charitable Foundation and the Keck, Snyder, Woo, and Yu Foundations, as well as by the National Institutes of Health, and the Defense Advanced Research Project Agency Reorganization and Plasticity to Accelerate Injury Recovery (DARPA REPAIR) program.

Author Contributions H.E.K. performed the structural determination of C1C2, prepared the mutants, measured the spectral property of C1C2 and wrote the paper. A.D.M. performed the electrophysiological analyses of C1C2. J.I. helped A.D.M. to take pictures of C1C2 and to determine membrane expression. Y.A. helped A.D.M. to perform patch-clamp experiments. T.T., T.N., R.I. and O.N. assisted with the structural determination. K.H. assisted with the data collection of C1C2. O.N., F.Z. and K.D. conceived the study; F.Z., O.Y. and K.D. helped to organize the project; S.H. and P.H. provided input on structural considerations; and C.R. with K.D. constructed the final C1C2 that enabled crystal structure determination. All authors discussed the results and commented on the manuscript. O.N. and K.D. supervised all aspects of the work and wrote/edited the manuscript.

Author Information Data have been deposited at the Protein Data Bank under accession number 3UG9. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to O.N. (nureki@biochem.s.u-tokyo.ac.jp) or K.D. (deissero@stanford.edu).

METHODS

Expression and purification of C1C2. Chimaeras between ChR1 and ChR2 from *C. reinhardtii* and other algal species were subcloned into the pCGFP-EU vector²¹ for expression in HEK293 cells. The tobacco etch virus (TEV) protease cleavage site, the coding sequence of enhanced GFP (EGFP), and the octa-histidine tag (EGFP-His₈) were introduced at the C terminus of the chimaeric constructs. All constructs were screened by FSEC analysis²¹. The gene encoding the best chimaera (C1C2) was subcloned into the modified pFastBac1 vector for expression in Sf+ insect cells. Baculovirus-infected Sf+ cells were cultured in Sf900II (Invitrogen) at 27 °C for 24 h, and then the temperature was reduced to 20 °C. Cells were harvested 72 h after infection by centrifugation at 6,000g for 10 min. The pellets were disrupted by two passages through a microfluidizer at 15,000 pounds per square inch, and were resuspended in a buffer containing 300 mM NaCl, 50 mM Tris-HCl, pH 8.0, 5% glycerol and 0.1 mM phenylmethylsulfonyl fluoride (PMSF). The cell debris was cleared by centrifugation at 10,000g for 40 min, and the crude membrane fraction was collected by ultracentrifugation (Ti45 rotor, 43,000 r.p.m., 1 h). This fraction was solubilized in a buffer containing 300 mM NaCl, 50 mM Tris-HCl, pH 8.0, 5% glycerol, 20 mM imidazole, 0.1 mM PMSF, 2.5% *n*-dodecyl- β -D-maltoside (DDM) and 0.5% cholesteryl hemisuccinate (CHS). The insoluble material was removed by ultracentrifugation (Ti70 rotor, 45,000 r.p.m., 30 min), and the supernatant was mixed with Ni-NTA resin (QIAGEN). After binding for 1 h, C1C2 was eluted in buffer supplemented with 300 mM imidazole. Following the cleavage of EGFP-His₈ by His-tagged TEV protease (homemade), the sample was reloaded onto the Ni-NTA column to remove the cleaved EGFP-His₈. The flow-through containing C1C2 was collected, concentrated, and further purified by size-exclusion chromatography in 150 mM NaCl, 50 mM Tris-HCl, pH 8.0,

5% glycerol, 0.05% DDM and 0.01% CHS. Peak fractions were pooled and concentrated to $\sim 10 \text{ mg ml}^{-1}$ for crystallization. For the mercury derivative, the concentrated protein was incubated with a sixfold molar excess of methyl mercury chloride at 20 °C for 1 h. The derivative was ultracentrifuged and used for crystallization experiments.

Crystallization. C1C2 was mixed with monoolein (Sigma) in a 2:3 protein to lipid ratio (w/w). Aliquots (100 nl) of the protein-LCP mixture were spotted on a 96-well sandwich plate and overlaid by 1 μ l of precipitant solution by the crystallization robot, mosquito LCP (TTP LabTech). Native crystals were obtained in 30–34% (w/v) PEG500DME, 100 mM Na citrate, pH 6.0, 100 mM MgCl₂, 100 mM NaCl and 100 mM (NH₄)₂SO₄, whereas the derivative crystals were grown in 31% (w/v) PEG500DME, 100 mM HEPES-NaOH, pH 7.0, 200 mM Li₂SO₄ and 10 mM ATP. All crystals were incubated for 2–3 weeks in the dark. They were harvested using micromounts (MiTeGen), and were flash-cooled in liquid nitrogen without any additional cryoprotectant.

Structure determination. X-ray diffraction data sets for the native and mercury-derivatized protein crystals were collected on beamline X06SA at SLS and beamline BL32XU at SPring-8, using a 1- μ m-wide, 15- μ m-high microbeam³⁸. Data were indexed and scaled with the programs XDS³⁹ and SCALA⁴⁰, or with DENZO and SCALEPACK from the HKL2000 program suite (HKL Research). Experimental phases were determined by the MAD method, using the four Hg sites identified with the program SHELX⁴¹. Subsequent refinements of the heavy atom parameters and phase calculations were performed with the program SHARP⁴². The data collection and phasing statistics are shown in Supplementary Table 1. The initial model structure of C1C2 was built with the program Phyre⁴³, using the *Anabaena* sensory rhodopsin structure (PDB accession 1XIO) as the template. The resultant structure was manually modified to fit into the experimental electron density maps, using the program Coot⁴⁴. The structure was

then refined with the program Phenix⁴⁵. Figures were prepared with Cuemol (<http://www.cuemol.org>).

Electrophysiology. HEK293 cells were cultured on poly-lysine-coated, glass-bottom culture dishes (Matsunami), and were transfected with 0.5 μ g of a plasmid construct containing the GFP-tagged C1C2 or the GFP-tagged C1C2 mutants. At 24–30 h after transfection, the cells were placed in a bath medium, containing 140 mM NaCl, 1 mM CaCl₂, 2 mM MgCl₂, 10 mM HEPES and 5 mM glucose (pH 7.4 with NaOH), under an inverted microscope (Olympus IX71). Calcium and potassium photocurrents were measured by replacing 140 mM NaCl by 90 mM CaCl₂ or 140 mM KCl accordingly. For proton photocurrents, cell bath was 5 mM NaCl, 135 mM *N*-methyl-D-glucamine, 1 mM CaCl₂, 2 mM MgCl₂, 10 mM HEPES and 5 mM glucose (pH 6.4). A borosilicate patch pipette (Harvard Apparatus), with a resistance of about 5–8 M Ω , was filled with 140 mM KCl, 5 mM EGTA, 2 mM MgCl₂ and 10 mM HEPES (pH 7.2 with KOH). C1C2 currents were recorded in the voltage-clamp mode and in the whole-cell configuration. The cells were held at a membrane potential of -80 mV , and were depolarized by 10 mV voltage steps of 1.8 s up to $+70 \text{ mV}$. The light-dependent currents were activated 200 ms after the depolarization step, with 465 nm light (1.5 mW mm^{-2}) for 1,000 ms, elicited by a high power LED illumination system (LEX2-B, Brainvision) connected to an A/D converter (Digidata 1440, Axon CNS, Molecular Devices), controlled by the pClamp10 software (Axon CNS). The light power was 1.5 mW mm^{-2} . Currents were measured using an Axopatch 200B amplifier (Axon CNS, Molecular Devices), filtered at 2 KHz, and sampled at 5 KHz, using a Digidata 1440A digitizer (Axon CNS) controlled by the pClamp10 software (Axon CNS).

Fluorescence measurements. The cells were transfected with 0.5 μ g wild-type C1C2 or C1C2 mutants, using Fugene 6, for 30 h. The cells were then washed with PBS and fixed with 4% paraformaldehyde in PBS for 20 min at room temperature (20 °C), and washed again with PBS before microscopy observation. GFP fluorescence was observed with a laser confocal microscope (FV1000 Olympus). To estimate membrane expression of C1C2 and its mutants, the ratio between the membrane fluorescence and cytosolic were determined.

Ultraviolet/visible spectroscopy. Ultraviolet/visible absorption spectra were recorded with an Ultrospec 3300 pro ultraviolet/visible spectrophotometer (Amersham Biosciences) by use of 1-cm quartz cuvettes. Freshly prepared C1C2 was used for the measurements. pH was adjusted by addition of 100 volumes of buffer solution yielding final concentrations of 50 mM Na citrate, pH 4.0, 50 mM Na acetate, pH 5.0, 50 mM Na cacodylate, pH 6.0, 50 mM HEPES, pH 7.0, 50 mM Tris, pH 8.0 and 9.0, or 50 mM CAPS, pH 10.0, plus 150 mM NaCl, 5% glycerol, 0.05% DDM and 0.01% CHS.

38. Hirata, K. *et al.* New micro-beam beamline at SPring-8, targeting at protein microcrystallography. *AIIP Conference Proceedings* **1234**, 893–896 (2010).
39. Kabsch, W. XDS. *Acta Crystallogr. D* **66**, 125–132 (2010).
40. Evans, P. Scaling and assessment of data quality. *Acta Crystallogr. D* **62**, 72–82 (2006).
41. Schneider, T. R. & Sheldrick, G. M. Substructure solution with SHELXD. *Acta Crystallogr. D* **58**, 1772–1779 (2002).
42. de La Fortelle, E. & Bricogne, G. Maximumlikelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* **276**, 472–494 (1997).
43. Kelley, L. A. & Sternberg, M. J. Protein structure prediction on the Web: a case study using the Phyre server. *Nature Protocols* **4**, 363–371 (2009).
44. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
45. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).

Light echoes reveal an unexpectedly cool η Carinae during its nineteenth-century Great Eruption

A. Rest¹, J. L. Prieto^{2,3}, N. R. Walborn¹, N. Smith⁴, F. B. Bianco^{5,6}, R. Chornock⁷, D. L. Welch⁸, D. A. Howell^{5,6}, M. E. Huber⁹, R. J. Foley⁷, W. Fong⁷, B. Sinnott⁸, H. E. Bond¹, R. C. Smith¹⁰, I. Toledo¹¹, D. Minniti¹² & K. Mandel^{7,13}

η Carinae is one of the most massive binary stars in the Milky Way^{1,2}. It became the second-brightest star in our sky during its mid-nineteenth-century ‘Great Eruption’, but then faded from view (with only naked-eye estimates of brightness^{3,4}). Its eruption is unique in that it exceeded the Eddington luminosity limit for ten years. Because it is only 2.3 kiloparsecs away, spatially resolved studies of the nebula have constrained the ejected mass and velocity, indicating that during its nineteenth-century eruption, η Car ejected more than ten solar masses in an event that released ten per cent of the energy of a typical core-collapse supernova^{5,6}, without destroying the star. Here we report observations of light echoes of η Carinae from the 1838–1858 Great Eruption. Spectra of these light echoes show only absorption lines, which are blueshifted by -210 km s^{-1} , in good agreement with predicted expansion speeds⁶. The light-echo spectra correlate best with those of G2-to-G5 supergiants, which have effective temperatures of around 5,000 kelvin. In contrast to the class of extragalactic outbursts assumed to be analogues of the Great Eruption of η Carinae^{7–12}, the effective temperature of its outburst is significantly lower than that allowed by standard opaque wind models¹³. This indicates that other physical mechanisms such as an energetic blast wave may have triggered and influenced the eruption.

η -Car-like giant eruptions of luminous blue variable stars are characterized by significant mass-loss and an increase in luminosity by several magnitudes^{8,9}. It has been thought that this increase in luminosity drives a dense wind, producing an optically thick, cooler pseudo-photosphere with a minimum effective temperature of 7,000 K and an F-type spectrum¹³. Within this model, η Car has been considered the prototype of these “supernova imposters”^{7–12}.

We obtained images in proximity to η Car (Fig. 1) that, when differenced, show a rich set of light echoes. The largest interval between our images is eight years. We have also found similar echo candidates at other positions, which we are currently monitoring. The large brightening and long duration point to the Great Eruption as the source of the light echoes. We have also obtained a composite light curve in the Sloan Digital Sky Survey (SDSS) *i* filter of the light echoes (see Fig. 2), showing a slow decline of several tenths of a magnitude over half a year. This light curve is most consistent with the historical observations⁴ of a peak in 1843, part of the 1838–1858 Great Eruption, although further observations are necessary to be certain (see the Supplementary Information).

Spectra of the light echoes (see Fig. 3) show only absorption lines characteristic of cool stellar photospheres, but no evidence of emission lines. In particular, the Ca II infrared triplet is only observed as absorption lines in the spectrum. Because of bright ambient nebular emission, it is difficult to determine whether there is any H α emission from η Car

itself, but in any case it must be weak, if present. By cross-correlating each of our η Car echo spectra with the Ultraviolet and Visual Echelle Spectrograph (UVES) spectral library¹⁴ (see Supplementary Figs 2 and 3), we find the best agreement with supergiant spectral types in the range of G2 to G5, with an effective temperature of around 5,000 K. Spectral types of F7 or earlier are ruled out by our analysis (see Supplementary Information for more details).

Doppler shifts of absorption features in the echo spectra provide direct information about the outflow speeds during the eruption. The Ca II infrared triplet absorption features in the spectrum are noticeably blueshifted (see Supplementary Fig. 2). By cross-correlation with G-type¹⁵ templates, we determine velocities of -202 ± 9 , -210 ± 14 and $-237 \pm 17 \text{ km s}^{-1}$ (errors are standard deviation) for our three individual spectra and an average velocity of $-210 \pm 30 \text{ km s}^{-1}$, which includes an uncertainty for the dust sheet velocity.

The bipolar nature of the Homunculus nebula shows that the η Car Great Eruption was strongly aspherical. It has been predicted that the outflow speeds that one would derive from the spectra of η Car in outburst, looking at the poles and equator of the double lobes, would be about -650 km s^{-1} and -40 to -100 km s^{-1} , (note that the speeds are negative because they are blueshifted, that is, the outflow is moving towards us) respectively¹⁶ (outflow speeds near the equator have a steep latitude dependence). The light echo we investigate here arises from latitudes near the equator of η Car (see Supplementary Fig. 1), and the measured blueshifted velocity of $-210 \pm 30 \text{ km s}^{-1}$ is in good agreement with expansion speeds within $\pm 20^\circ$ of the equatorial plane. We also find a strong asymmetry in the Ca II infrared triplet, extending to a velocity of -850 km s^{-1} —well below the speed of the fastest polar ejecta found previously⁶, but in good agreement with speeds observed in the blast wave at lower latitudes⁶. Future observations of light echoes viewing the η Car eruption from different directions, in particular from the poles, has the potential to observe these very-high-velocity ejecta and other asymmetries.

A characteristic of luminous-blue-variable outbursts is their transition from a hot quiescent state to a cooler outburst state, although this feature is less well observed for the giant eruptions (see Fig. 4). Two potential models for luminous-blue-variable outbursts involve either an opaque stellar wind driven by an increase in luminosity, or a hydrodynamic explosion. The traditional mechanism for η -Car-like giant eruptions has been that an unexplained increase in luminosity drives a denser wind, so that an optically thick pseudo-photosphere forms at a layer much larger and cooler than the hydrostatic stellar surface¹³. This model predicts a minimum effective temperature of 7,000 K, resembling A- or F-type supergiants^{8,17}, because the wind opacity depends on the temperature (see Fig. 4). A giant eruption evidently occurs as a massive star attempts to evolve redward and

¹Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, Maryland 21218, USA. ²Carnegie Observatories, 813 Santa Barbara Street, Pasadena, California 91101, USA. ³Department of Astrophysical Sciences, Princeton University, Peyton Hall, Princeton, New Jersey 08544, USA. ⁴Steward Observatory, University of Arizona, 933 North Cherry Avenue, Tucson, Arizona 85721, USA. ⁵Las Cumbres Observatory Global Telescope Network, Goleta, California 93117, USA. ⁶Department of Physics, University of California, Santa Barbara, California 93106, USA. ⁷Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA. ⁸Department of Physics and Astronomy, McMaster University, Hamilton, Ontario, L8S 4M1, Canada. ⁹Department of Physics and Astronomy, Johns Hopkins University, Baltimore, 3400 North Charles Street, Maryland 21218, USA. ¹⁰Cerro Tololo Inter-American Observatory, National Optical Astronomy Observatory, Colina el Pino S/N, La Serena, Chile. ¹¹ALMA, KM 121 CH 23, San Pedro de Atacama, II Region, Chile. ¹²Department of Astronomy and Astrophysics, Pontificia Universidad Catolica de Chile, Santiago 22, Chile. ¹³Imperial College London, Blackett Laboratory, Prince Consort Road, London SW7 2AZ, UK.

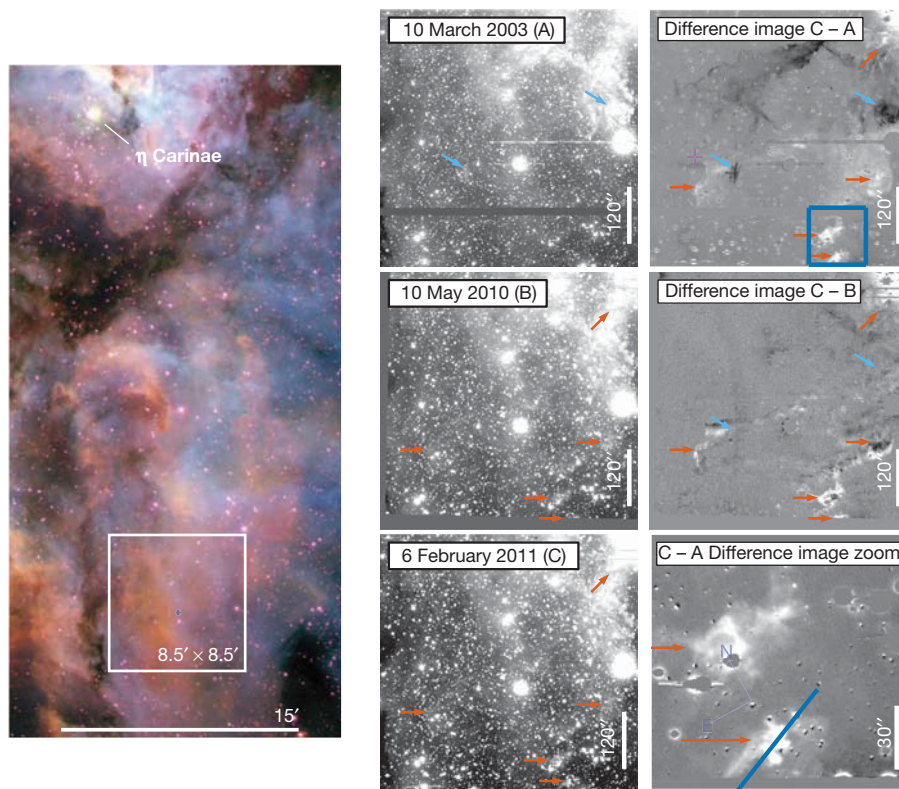


Figure 1 | η Car light echoes. The left panel shows the positions of the star η Carinae and our images (white box), plotted on an image in the light of three different emission lines: oxygen (blue), hydrogen (green) and sulphur (red). (Photo taken by N.S.) The middle panels show the images obtained with the CTIO 4-m Blanco telescope of a region about 0.5° to the south of η Car at 10 March 2003 (epoch A), 10 May 2010 (epoch B), and 6 February 2011 (epoch C), from top to bottom. The right panels show the difference images 'C minus A' and 'C minus B' at the top and middle, respectively. Example light-echo positions are indicated by blue (epoch A) and red (epochs B and C) arrows. The bottom right panel shows a zoom of the spectrograph slit, indicated with a

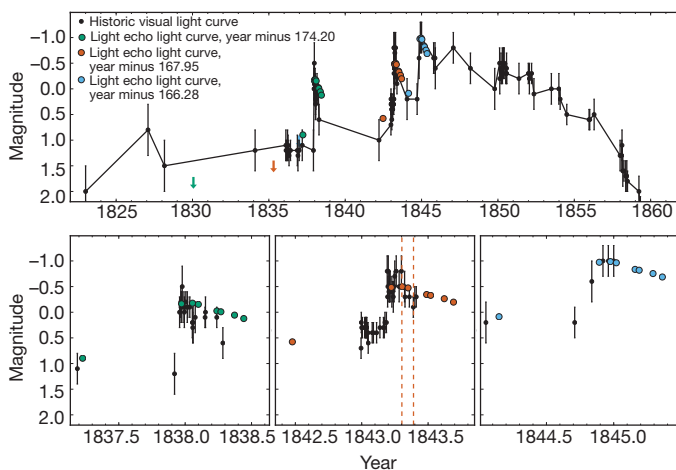


Figure 2 | Historical and light-echo lightcurve of η Car. The historical light curve⁴ in visual apparent magnitudes is shown with black circles and lines, with error bars indicating approximate uncertainties in these eye estimates. Light echo brightnesses (SDSS i ; error bars are the standard deviation) from our eight modern images spanning about eight years are displayed shifted by 174.2 Earth years (green circles), 167.95 years (red circles) and 166.28 years (blue circles), to illustrate the best-matching time delays for the 1838, 1843 and 1845 outbursts, respectively. The first epoch is an upper limit indicated with an arrow. The upper panel shows the full time range of the Great Eruption and therefore shows all three potential matches, whereas the lower panels show the brightnesses from seven of our eight modern epochs in a magnified time period around each peak.

blue line. For all panels north is up and east is to the left. Applying the vector method that previously allowed us to identify the source of the light echoes from the supernovae that produced the supernova remnants SNR 0509–67.5, Cas A and Tycho^{24,25}, we find that a dramatic brightening of η Car must be the origin. In these echoes, unlike those of Galactic supernovae²⁶, there is still significant spatial overlap even at separations of one light-year, suggesting that the duration of the event causing them must be significantly longer than one year. We also see brightening of two magnitudes or more within eight years. Thus, the Lesser Eruption from 1887 to 1896, which brightened by only a magnitude, is excluded as the source.

encounters the Humphreys–Davidson limit¹³, beyond which no stable stars are observed.

Surprisingly, our G-type light-echo spectrum of the η Car Great Eruption is inconsistent with expectations of an opaque-wind model¹³ (see Fig. 4). With this model, it is difficult to explain the high (10^{50} erg) kinetic energy⁵ and the presence of a fast blast wave at large radii⁶. Instead, these observations point towards a hydrodynamic explosion that must have influenced the Great Eruption^{2,5,6}.

The first visual spectroscopic observations of η Car around 1870 showed emission lines^{18,19}. A photographic spectrogram obtained during its Lesser Eruption^{20,21} around 1890 resembles an F-type supergiant blueshifted by -200 km s^{-1} , with moderate P Cygni hydrogen profiles, which is as expected in the opaque-wind model¹³. The difference between the 1890 spectrum and our light-echo spectrum of the Great Eruption is therefore quite striking, indicating that two distinct physical processes may have been involved for two outbursts of the same object. However, the 1890 event also produced a mass ejection, the Little Homunculus, with the same axial symmetry (although much smaller mass) as the Great Eruption²².

Luminous-blue-variable giant eruptions are rare, and have only been recorded twice in our Galaxy in the last 400 years: the Great Eruption of η Car and the giant eruption in the seventeenth century of P Cygni. Because of their considerable intrinsic brightness just below the luminosity of faint core-collapse supernovae, about two dozen giant eruption candidates, called supernova imposters because they have often been mistaken for supernovae, have been found in various extragalactic supernova searches^{7–12}. Typically, the hotter

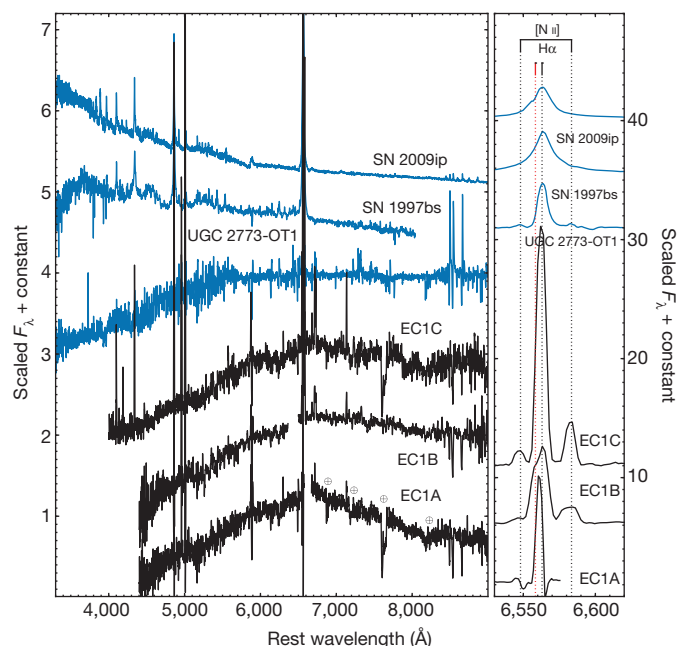


Figure 3 | Light echo spectra of the Great Eruption of η Car. The three optical low-resolution spectra of the light echo (black lines) were taken at J2000 position right ascension 10 h 44 min 12.127 s and declination $-60^{\circ} 16' 01.69''$ in March and April 2011 obtained at the Magellan I 6.5-m and du Pont 2.5-m telescopes of the Las Campanas Observatory, Chile. A log of the spectroscopic observations and details of the spectra is presented in Supplementary Table 1. The slit positions differ only slightly in slit angle. The spectra were reduced using standard techniques and then wavelength-calibrated using observations of an HeNeAr lamp. The wavelength calibration was checked and corrected using night-sky emission lines, especially [O I] $\lambda 5,577$ Å, and OH lines in the red part of the spectrum. We flux-calibrated the spectra using a flux standard observed the same night as the science observations. The left panel shows the spectra from 5,000 to 9,000 Å. The spectra are not corrected for reddening nor for the blueward scattering by the dust. For comparison, the blue lines show spectra of three examples of supernova imposters: SN 1997bs, SN 2009ip and UGC 2773-OT1. The right panel shows the H α and [N II] emission lines. We note that the background emission-line subtraction is incomplete because the emission lines vary spatially. Also, EC1A H α is at the edge of the chip and is therefore uncertain. Crossed circles indicate the locations of atmospheric absorption lines.

supernova imposters have steep blue continua, stronger and broader Balmer lines, and relatively weak absorption, whereas the cooler ones tend to have redder continua, weaker and narrower Balmer lines, strong [Ca II] and Ca II emission, deeper P Cygni absorption features, and in some cases stronger absorption spectra similar to those of F-type supergiants²³. However, the η Car Great Eruption light-echo spectrum is quite different. Its spectral type is G2 to G5, significantly later than all other supernova imposters at peak. Furthermore, the Ca II infrared triplet lines are only in absorption. For the extreme mass-loss rates required for the Great Eruption of η Car, another process must give rise to the apparent temperature.

The Great Eruption of η Car has been considered the prototype of the extragalactic supernova imposters or η Car analogues, even though it is actually an extreme case in terms of radiated energy ($10^{49.3}$ erg), kinetic energy ($>10^{50}$ erg), and its decade-long duration²³. The spectra of the light echo indicates now that it is not only extreme, but a different, unique object. It is difficult to see how strong emission lines could be avoided in an opaque wind where the continuum photosphere is determined by a change in opacity, and its temperature and broad absorption lines are more consistent with the opaque cooling photosphere of an explosion. What triggered such an explosion and the reason that the huge mass-loss did not destroy the star are still unknown, but predictions from future radiative transfer simulations trying to explain η Car and its Great Eruption can now be matched to

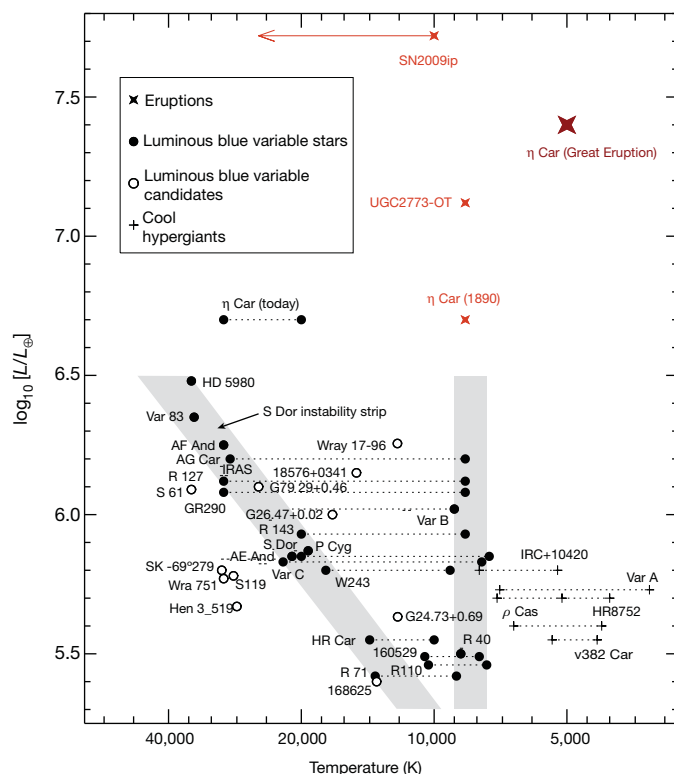


Figure 4 | Hertzsprung–Russell diagram with luminous blue variables and η Car. Adaptation¹⁷ of a Hertzsprung–Russell diagram showing luminous blue variables, related hypergiant stars, and the peak luminosities of luminous-blue-variable-like eruptions. The grey bands denote the typical locations of luminous blue variables in quiescence (left, diagonal band) and during the S Doradus-like outburst. Temperatures for the Great Eruption and the 1890 eruption of η Car are based on the echo spectra presented here and the F-type spectrum of the 1890 event²⁰, respectively. The temperature of 10,000 K for SN 2009ip is based on the observed continuum shape, but this is only a lower limit because of the possible effects of circumstellar or host galaxy reddening²⁷. Because of the presence of He I lines in the spectrum, the true temperature is probably much hotter. The 8,500 K temperature of UGC2773-OT is indicated by the F-type absorption features in its spectrum, and this temperature is relatively independent of reddening^{27,28}.

these spectral observations. Alternative models, such as the ones that use mass accretion from the companion star during periastron passage as a trigger for the eruption¹², can be verified or dismissed.

Received 26 August; accepted 8 December 2011.

- Davidson, K. & Humphreys, R. M. Eta Carinae and its environment. *Annu. Rev. Astron. Astrophys.* **35**, 1–32 (1997).
- Damineli, A. The 5.52 year cycle of eta Carinae. *Astrophys. J.* **460**, L49 (1996).
- Frew, D. J. The historical record of η Carinae I. The visual light curve, 1595–2000. *J. Astron. Data* **10**, 6 (2004).
- Smith, N. & Frew, D. J. A revised historical light curve of eta Carinae and the timing of close periastron encounters. *Mon. Not. R. Astron. Soc.* **415**, 2009–2019 (2011).
- Smith, N. *et al.* Mass and kinetic energy of the Homunculus nebula around η Carinae. *Astron. J.* **125**, 1458–1466 (2003).
- Smith, N. A blast wave from the 1843 eruption of η Carinae. *Nature* **455**, 201–203 (2008).
- Goodrich, R. W., Stringfellow, G. S., Penrod, G. D. & Filippenko, A. V. S. N. 1961V—an extragalactic ETA Carinae analog. *Astrophys. J.* **342**, 908–916 (1989).
- Humphreys, R. M. & Davidson, K. The luminous blue variables: astrophysical geysers. *Publ. Astron. Soc. Pacif.* **106**, 1025–1051 (1994).
- Humphreys, R. M., Davidson, K. & Smith, N. Eta Carinae's second eruption and the light curves of the eta Carinae variables. *Publ. Astron. Soc. Pacif.* **111**, 1124–1131 (1999).
- Van Dyk, S. D. *et al.* SN 1997bs in M66: another extragalactic η Carinae analog? *Publ. Astron. Soc. Pacif.* **112**, 1532–1541 (2000).
- Vink, J. S. Eta Carinae and the luminous blue variables. Preprint at <http://arxiv.org/abs/0905.3338> (2009).
- Kashi, A., Frankowski, A. & Soker, N. NGC 300 OT2008–1 as a scaled down version of the eta Carinae great eruption. *Astrophys. J.* **709**, L11–L15 (2010).

13. Davidson, K. The relation between apparent temperature and mass-loss rate in hypergiant eruptions. *Astrophys. J.* **317**, 760–764 (1987).
14. Bagnulo, S. *et al.* The UVES Paranal Observatory Project: a library of high-resolution spectra of stars across the Hertzsprung-Russell diagram. *The Messenger* **114**, 10–14 (2003).
15. Cenarro, A. J. *et al.* Empirical calibration of the near-infrared Ca II triplet—I. The stellar library and index definition. *Mon. Not. R. Astron. Soc.* **326**, 959–980 (2001).
16. Smith, N. The structure of the Homunculus. I. Shape and latitude dependence from H_2 and [Fe II] velocity maps of η Carinae. *Astrophys. J.* **644**, 1151–1163 (2006).
17. Smith, N., Vink, J. S. & de Koter, A. The missing luminous blue variables and the bistability jump. *Astrophys. J.* **615**, 475–484 (2004).
18. Le Sueur, A. On the nebulae of Argo and Orion, and on the spectrum of Jupiter. *Proc. R. Soc. Lond. I* **18**, 245–250 (1869).
19. Le Sueur, A. Observations with the Great Melbourne telescope, in a letter to Prof. Stokes. *Proc. R. Soc. Lond. I* **19**, 18–19 (1870).
20. Walborn, N. R. & Liller, M. H. The earliest spectroscopic observations of η Carinae and its interaction with the Carina nebula. *Astrophys. J.* **211**, 181–183 (1977).
21. Humphreys, R. M., Davidson, K. & Koppelman, M. The early spectra of η Carinae 1892 to 1941 and the onset of its high excitation emission spectrum. *Astron. J.* **135**, 1249–1263 (2008).
22. Smith, N. Doppler tomography of the Little Homunculus: high-resolution spectra of [Fe II] λ 16435 around η Carinae*. *Mon. Not. R. Astron. Soc.* **357**, 1330–1336 (2005).
23. Smith, N., Li, W., Silverman, J. M., Ganeshalingam, M. & Filippenko, A. V. Luminous blue variable eruptions and related transients: diversity of progenitors and outburst properties. *Mon. Not. R. Astron. Soc.* **415**, 773–810 (2011).
24. Rest, A. *et al.* Light echoes from ancient supernovae in the Large Magellanic Cloud. *Nature* **438**, 1132–1134 (2005).
25. Rest, A. *et al.* Scattered-light echoes from the historical galactic supernovae Cassiopeia A and Tycho (SN 1572). *Astrophys. J.* **681**, L81–L84 (2008).
26. Rest, A. *et al.* On the interpretation of supernova light echo profiles and spectra. *Astrophys. J.* **732**, 2 (2011).
27. Smith, N. *et al.* Discovery of precursor luminous blue variable outbursts in two recent optical transients: the fitfully variable missing links UGC 2773-OT and SN 2009ip. *Astron. J.* **139**, 1451–1467 (2010).
28. Foley, R. J. *et al.* The diversity of massive star outbursts. I. Observations of SN2009ip, UGC 2773 OT2009–1, and their progenitors. *Astrophys. J.* **732**, 32 (2011).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. Humphreys, K. Davidson and J. Vink for comments and discussions. We thank S. Blondin for help with the continuum subtraction. The Blanco 4-m telescope is a facility of the Cerro Tololo Inter-American Observatory, National Optical Astronomy Observatory, which is operated by the Association of Universities for Research in Astronomy, under contract with the National Science Foundation. We use data from the UVES Paranal Observatory Project. The computations in this paper were run on the Odyssey cluster supported by the FAS Science Division Research Computing Group at Harvard University. Observations were obtained at LCOGT, and F.B.B. and D.A.H. acknowledge support from LCOGT. J.L.P. is a Hubble Carnegie-Princeton Fellow. R.J.F. is a Clay Fellow.

Author Contributions All authors contributed to the drafting of the paper. A.R., N.S. and R.C.S. imaged the area around η Car. A.R. and M.E.H. reduced the imaging data. H.E.B. provided images of the echoes that guided our spectroscopic pointings. J.L.P., R.C., R.J.F. and W.F. obtained the spectra and reduced them. A.R. and J.L.P. performed spectral analysis and interpretation. A.R., N.R.W. and F.B.B. performed spectral classification. F.B.B. and K.M. correlated the spectra. A.R., D.L.W. and B.S. modelled the light echo. I.T. and D.M. provided imaging of η Car. F.B.B. and D.A.H. provided the FTS images, and F.B.B. and A.R. reduced them.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.R. (arest@stsci.edu).

Magnetic reconnection from a multiscale instability cascade

Auna L. Moser¹ & Paul M. Bellan¹

Magnetic reconnection, the process whereby magnetic field lines break and then reconnect to form a different topology, underlies critical dynamics of magnetically confined plasmas in both nature^{1–4} and the laboratory^{5–9}. Magnetic reconnection involves localized diffusion of the magnetic field across plasma, yet observed reconnection rates are typically much higher than can be accounted for using classical electrical resistivity¹⁰. It is generally proposed¹⁰ that the field diffusion underlying fast reconnection results instead from some combination of non-magnetohydrodynamic processes that become important on the ‘microscopic’ scale of the ion Larmor radius or the ion skin depth. A recent laboratory experiment¹¹ demonstrated a transition from slow to fast magnetic reconnection when a current channel narrowed to a microscopic scale, but did not address how a macroscopic magnetohydrodynamic system accesses the microscale. Recent theoretical models¹² and numerical simulations^{13,14} suggest that a macroscopic, two-dimensional magnetohydrodynamic current sheet might do this through a sequence of repetitive tearing and thinning into two-dimensional magnetized plasma structures having successively finer scales. Here we report observations demonstrating a cascade of instabilities from a distinct, macroscopic-scale magnetohydrodynamic instability to a distinct, microscopic-scale (ion skin depth) instability associated with fast magnetic reconnection. These observations resolve the full three-dimensional dynamics and give insight into the frequently impulsive nature of reconnection in space and laboratory plasmas.

The experiment (Fig. 1) involves a long, slender, current-carrying magnetized plasma jet that evolves over the course of $\sim 50 \mu\text{s}$. The jet front travels at $\sim 10 \text{ km s}^{-1}$, increasing the jet length until the current-driven kink instability, an ideal magnetohydrodynamic (MHD) phenomenon, sets in¹⁵ and deforms the plasma jet into a helical structure the amplitude of which grows in time (Fig. 2, Supplementary Fig. 1 and Supplementary Movie 1). The experiments used hydrogen, nitrogen or argon plasma. In all three of these, the kink amplitude growth rate was observed to have one of two distinct behaviours: linear growth or exponential growth (Fig. 2b). Because argon provides the clearest images, argon data will be used in the following detailed discussion.

In the case of an exponentially growing kink, the jet segment develops a periodic fine structure (Fig. 3). The fine-structure growth rate, location and spatial periodicity are consistent with the magnetized plasma Rayleigh–Taylor instability¹⁶. This instability develops in a gravitational field at an interface where a heavy fluid with density ρ_2 lies above a light fluid with density ρ_1 . The acceleration of the exponentially growing kink segment creates an effective gravitational field, g_{eff} , in the plasma frame. To an observer in the frame of the accelerating filamentary kink segment at the location of the periodic structure (on the inward side of the outward accelerating filament, that is, the trailing side), the plasma filament appears to be a heavy fluid sitting on top of trailing low-density fluid immediately exterior to the filament.

The fastest-growing mode of the magnetized plasma Rayleigh–Taylor instability has $\mathbf{k} \cdot \mathbf{B} = 0$ (where \mathbf{k} is the instability wavevector and \mathbf{B} is the magnetic field vector), a property that will be useful in the

interpretation of the experiment. If we assume that the plasma density inside the filament greatly exceeds the density immediately outside, that is, $\rho_2 \gg \rho_1$, then the exponential growth rate for the fastest-growing Rayleigh–Taylor mode is¹⁶ $\gamma_{\text{RT}} \approx \sqrt{g_{\text{eff}} k}$, where $k = 2\pi/\lambda$ and λ is the wavelength of the fine structure. For the plasma shown in Fig. 3, the calculated growth rate is $\gamma_{\text{RT}} \approx 3 \times 10^6 \text{ m s}^{-1}$ and the visually observed

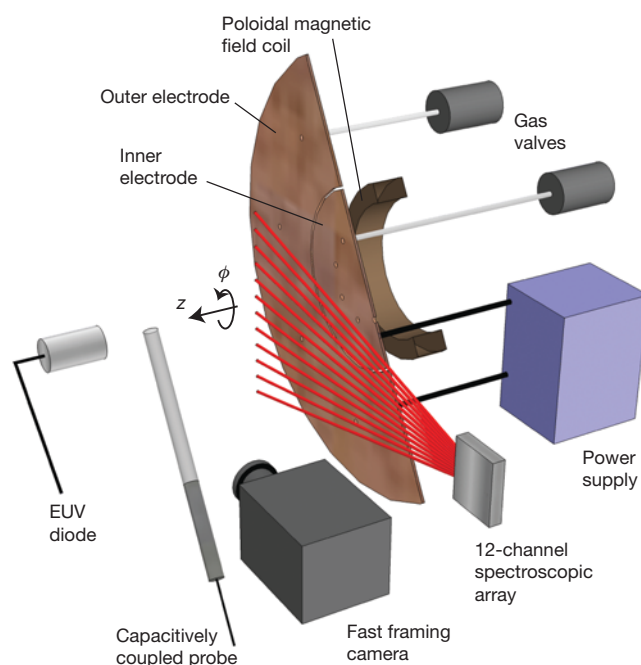
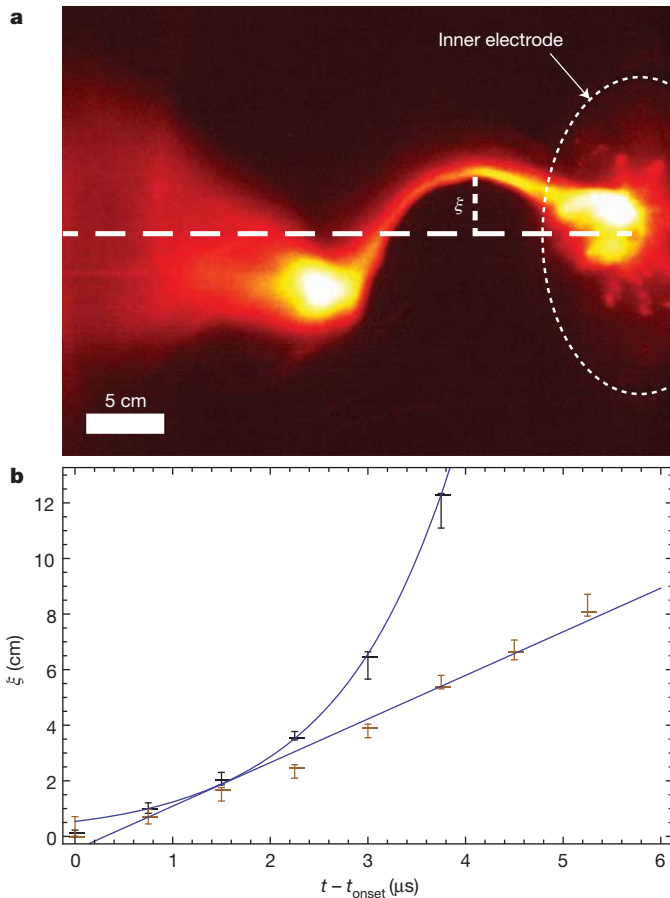


Figure 1 | Geometry of experimental set-up. The experiment produces a magnetized plasma jet using the coplanar disk and annulus electrodes and the background poloidal magnetic field coil shown in cut-away view. The 20-cm-diameter cathode (inner electrode) and the surrounding 50-cm-outer diameter anode (outer electrode) are mounted on the end dome of a 1.4-m-diameter, 1.6-m-long stainless steel vacuum chamber. To make the magnetized plasma jet, first the coil creates a poloidal magnetic field linking the disk to the annulus. Fast pulsed valves then puff in gas at eight locations in front of the anode and eight locations in front of the cathode. A high-energy capacitor bank then applies a potential of 5–6 kV across the electrodes, breaking the gas down into a plasma. The electric current from the capacitor bank ramps up to a peak amplitude of $\sim 110 \text{ kA}$, with a full-width at half-maximum of $\sim 60 \mu\text{s}$. The electric current flows along the jet column, producing a $\sim 0.1\text{-T}$ toroidal (ϕ direction) magnetic field, and completes its path along an outer, low-density, low-poloidal-magnetic-field shroud surrounding the jet. The jet is propelled in the z direction by the combination of the axial gradient of the toroidal magnetic field and the axial gradient of the hydrodynamic pressure resulting from the radial pinch force due to the current²². Diagnostic devices include a visible-light fast framing camera with a 20-ns shutter and an adjustable $\sim 1\text{-}\mu\text{s}$ interframe time, a gated, 12-channel linear spectroscopic array²³ (lines of sight indicated in red) with a 2-cm line-of-sight separation and a $1\text{-}\mu\text{s}$ time resolution, a capacitively coupled probe and extreme-ultraviolet (EUV) diodes sensitive within the 10–75-eV energy range.

¹Applied Physics, California Institute of Technology, Pasadena, California 91125, USA.



growth rate is $\gamma_{RT} \approx 10^6 \text{ m s}^{-1}$. The agreement between these two rates and the observation that the instability is located on the trailing side of the transversely accelerated filament together support the conclusion that the fine-scale instability is a Rayleigh–Taylor instability.

Figure 2 | Growth of the amplitude of the kink instability. **a**, Example of measurement of the kink amplitude, ξ , in a fast framing camera image of shot no. 10,934. The plasma jet undergoes a kink instability in agreement with the prediction of Kruskal–Shafranov theory^{15,24}; the jet is unstable to kinking when $\mu_0 I / \Psi > 4\pi/L$, where L is the jet length, I is the current measured at the electrodes and Ψ is the poloidal magnetic flux measured at the electrodes. Image intensity is logarithmically scaled and false coloured. **b**, Comparison of exponential kink growth rate and linear kink growth rate as measured from fast camera images. Black points indicate kink amplitude with exponential growth, $\xi \propto \exp(\gamma_{\text{kink}} t)$, where $\gamma_{\text{kink}} = 8.3 \times 10^5 \text{ s}^{-1}$ (shot no. 10,934). Brown points indicate kink amplitude with linear growth, $\xi \propto t$ (shot no. 10,930). Lines show best fits to the data. Error bars indicate the range of possible plasma edge positions based on image intensity.

When the thin filament breaks up as a result of the Rayleigh–Taylor instability, the portion of the jet beyond the break-up region retains its magnetic structure and separates from the remaining jet base (Fig. 3 and Supplementary Fig. 2), demonstrating a clear magnetic reconnection. Several additional diagnostics support the conclusion that magnetic reconnection is occurring. Photodiodes sensitive within the 10–75-eV energy range measure a burst of extreme ultraviolet radiation coincident with the filament break-up (Supplementary Fig. 3). A capacitively coupled probe placed in the plasma jet between the electrodes and the filament measures an order of magnitude increase in emissions in the whistler-wave frequency range coincident with the filament break-up¹⁷ (Supplementary Fig. 4).

We can verify that the reconnecting plasma is indeed at the non-MHD microscale by recalling that ideal MHD is based on the presumption that $v_d/v_A \ll 1$. Here $v_d = J_z/nq$ is the electron drift velocity for an axial current density J_z , where n is particle density and q is electron charge; and $v_A = B/\sqrt{\mu_0 n m_i}$ is the Alfvén velocity, where B is the magnetic field strength, m_i is the ion mass and μ_0 is the permeability of free space. When v_d/v_A becomes of order one, the assumptions underlying ideal MHD fail because kinetic effects (that is, wave–particle interaction) and Hall term effects (that is, decoupling of electron and ion perpendicular motions) become important. If we assume that $B \approx B_z$ and use $\mathbf{k} \cdot \mathbf{B} = 0$ (which holds for the fastest-growing Rayleigh–Taylor mode), we see that

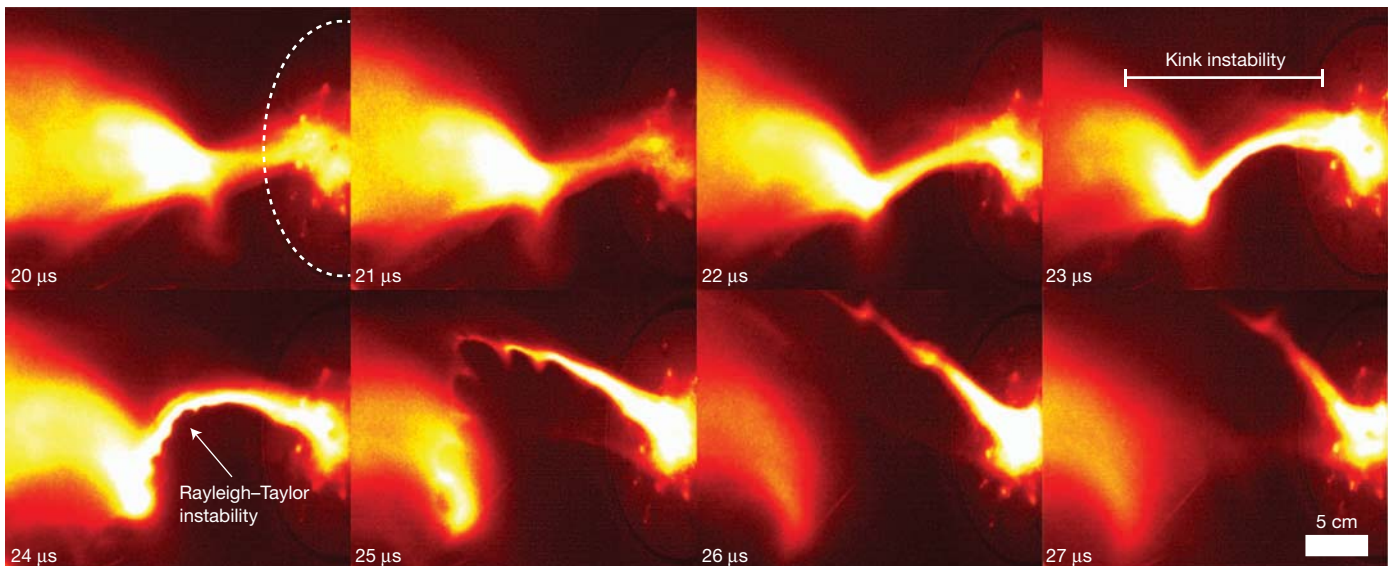


Figure 3 | Time series images of plasma jet evolution. For exponential kink amplitude growth, shown here, a segment of the kinked jet quickly narrows to a thin filament, which then brightens while developing a sharp, distinctive, periodic fine structure on the trailing side of the radially outward-accelerating filament. As the fine-structure amplitude grows, it erodes the filament diameter until the filament breaks up. The dashed line in the first image shows the position of the inner electrode. The measured transverse acceleration of the filament when the periodic fine structure first appears is $g_{\text{eff}} \approx 4 \times 10^{10} \text{ m s}^{-2}$.

The axial wavelength, as measured from the fast camera images, is $\lambda_z \approx 2 \text{ cm}$, which implies that $k \approx 300 \text{ m}^{-1}$. These measurements give a calculated Rayleigh–Taylor growth rate of $\gamma_{RT} \approx 3 \times 10^6 \text{ s}^{-1}$. For comparison, the growth rate can be estimated directly from the fast camera images. The fine-structure amplitude is $\sim 0.5 \text{ cm}$ when it first appears and grows to $\sim 1.5 \text{ cm}$ in $1 \mu\text{s}$. This corresponds to an observed growth rate of $\gamma_{RT} \approx 1 \times 10^6 \text{ s}^{-1}$, in agreement with the calculated growth rate. All images are from shot no. 11,225; image intensity is logarithmically scaled and false coloured. See also Supplementary Movie 2.

$$\frac{\mu_0 J_z}{B_z} = \frac{1}{r} \frac{\partial}{\partial r} (r B_\phi) \frac{1}{B_z} = \frac{1}{r} \frac{\partial}{\partial r} \left(r^2 2\pi \frac{B_z}{\lambda_z} \right) \frac{1}{B_z}$$

where λ_z is the local axial wavelength. This means that $v_d/v_A = (4\pi/\lambda_z)(c/\omega_{pi})$, where c is the speed of light and ω_{pi} is the ion plasma frequency (c/ω_{pi} is the ion skin depth).

Using c/ω_{pi} calculated with $n \approx 10^{22} \text{ m}^{-3}$ as determined from Stark-broadening density measurements and the measured fine-structure local axial wavelength, $\lambda_z = 2 \text{ cm}$, we find that $v_d/v_A \approx O(1)$. This shows that the filament observed to have a Rayleigh–Taylor instability is in the non-MHD regime at the time of magnetic reconnection.

Because the original kink instability was in the ideal-MHD regime, our argon plasma observations thus show a cascade from an ideal-MHD macroscopic kink instability to a non-ideal, microscopic Rayleigh–Taylor instability associated with magnetic reconnection. A question that remains to be answered is what determines whether the kink instability grows linearly or exponentially.

By contrast with its behaviour in argon, the Rayleigh–Taylor instability in hydrogen fails to break the plasma filament. Although we still observe an exponentially growing kink and subsequent Rayleigh–Taylor instability in hydrogen, the ion skin depth, c/ω_{pi} , is substantially smaller than in argon and the Rayleigh–Taylor instability fails to erode the plasma filament to a diameter less than this smaller depth (Supplementary Fig. 5), though we are not yet certain why. Thus, the hydrogen plasma diameter has not been reduced to the necessary microscale and no magnetic reconnection takes place.

Reconnection is observed in the nominal parameter regime of these experiments only when preceded by the Rayleigh–Taylor instability; this implies that the Rayleigh–Taylor instability is necessary for the observed reconnection to occur. However, the comparison between argon and hydrogen plasmas shows that the mere existence of the Rayleigh–Taylor instability is not sufficient for reconnection: the Rayleigh–Taylor instability must become large enough to erode the filament diameter to less than the ion skin depth for magnetic reconnection to occur.

Our observations demonstrate one possible mechanism by which a macroscopic MHD system can couple to the microscale processes necessary for magnetic reconnection. The essential components of this mechanism have been separately observed in nature. For example, in the solar corona current-carrying magnetic flux tubes confining plasma with density higher than the ambient value are common, as is kinking of such flux tubes^{18,19}, and Rayleigh–Taylor instabilities have been observed²⁰.

Given that we observe this mechanism in a range of laboratory experiments, we think it quite plausible that it will also occur in astrophysical systems with appropriate physical parameters. As a possible example, we note that one solar observation²¹ reported the lateral acceleration of a plasma-filled flux tube at $\sim 1 \text{ km s}^{-2}$. Assuming that such acceleration is typical, for a 10^4 -km-wide solar loop a Rayleigh–Taylor disturbance with an axial wavelength of 400 km and an initial amplitude of 0.5 km would grow exponentially to the 10^4 -km loop width in about 1 min and so erode the current channel to a microscopic scale. Because the instability wavelength assumed in this example is at the margin of existing

resolution capabilities, measurements that simultaneously resolve the widely separated macro- and microscales would be challenging.

Received 27 July 2011; accepted 3 January 2012.

1. Paschmann, G. *et al.* Plasma acceleration at the Earth's magnetopause: evidence for reconnection. *Nature* **282**, 243–246 (1979).
2. Mikic, Z., Barnes, D. C. & Schnack, D. D. Dynamical evolution of a solar coronal magnetic field arcade. *Astrophys. J.* **328**, 830–847 (1988).
3. Colgate, S., Li, H. & Pariev, V. The origin of the magnetic fields of the universe: the plasma astrophysics of the free energy of the universe. *Phys. Plasmas* **8**, 2425–2431 (2001).
4. Uzdensky, D. A. & MacFadyen, A. I. Stellar explosions by magnetic towers. *Astrophys. J.* **647**, 1192–1212 (2006).
5. Wesson, J. A. Sawtooth reconnection. *Nucl. Fusion* **30**, 2545–2549 (1990).
6. Yamada, M. *et al.* Investigation of magnetic reconnection in a high-temperature tokamak plasma. *Phys. Plasmas* **1**, 3269–3276 (1994).
7. Bellan, P. M. *Spheromaks* 60–75 (Imperial Coll. Press, 2000).
8. Taylor, J. B. Relaxation and magnetic reconnection. *Rev. Mod. Phys.* **58**, 741–763 (1986).
9. Brown, M. Experimental studies of magnetic reconnection. *Phys. Plasmas* **6**, 1717–1724 (1999).
10. Yamada, M., Kulsrud, R. & Ji, H. Magnetic reconnection. *Rev. Mod. Phys.* **82**, 603–664 (2010).
11. Egedal, J. *et al.* Laboratory observations of spontaneous magnetic reconnection. *Phys. Rev. Lett.* **98**, 015003 (2007).
12. Shibata, K. & Tanuma, S. Plasmoid-induced-reconnection and fractal reconnection. *Earth Planets Space* **53**, 473–482 (2001).
13. Che, H., Drake, J. F. & Swisdak, M. A current filamentation mechanism for breaking magnetic field lines during reconnection. *Nature* **474**, 184–187 (2011).
14. Drake, J. F. *et al.* Formation of secondary islands during magnetic reconnection. *Geophys. Res. Lett.* **303**, L13105 (2006).
15. Hsu, S. C. & Bellan, P. M. Experimental identification of the kink instability as a poloidal flux amplification mechanism for coaxial gun spheromak formation. *Phys. Rev. Lett.* **90**, 215002 (2003).
16. Chandrasekhar, S. *Hydrodynamic and Hydromagnetic Stability* 464–466 (Dover, 1961).
17. Gekelman, W. & Stenzel, R. L. Magnetic field line reconnection experiments: 6. Magnetic turbulence. *J. Geophys. Res.* **89**, 2715–2733 (1984).
18. Srivastava, A. K., Zaqarashvili, T. V., Kumar, P. & Khoachenko, M. L. Observation of kink instability during small B5.0 solar flare on 2007 June 4. *Astrophys. J.* **715**, 292–299 (2010).
19. Zhou, G. P. *et al.* Two successive coronal mass ejections driven by the kink and drainage instabilities of an eruptive prominence. *Astrophys. J.* **651**, 1238–1244 (2006).
20. Berger, T. *et al.* Magneto-thermal convection in solar prominences. *Nature* **472**, 197–200 (2011).
21. Liu, C. *et al.* The eruption from a sigmoidal solar active region on 2005 May 13. *Astrophys. J.* **669**, 1372–1381 (2007).
22. Kumar, D. & Bellan, P. M. Nonequilibrium Alfvénic plasma jets associated with spheromak formation. *Phys. Rev. Lett.* **103**, 105003 (2009).
23. Yun, G. S. & Bellan, P. M. Plasma tubes becoming collimated as a result of magnetohydrodynamic pumping. *Phys. Plasmas* **17**, 062108 (2010).
24. Hsu, S. C. & Bellan, P. M. A laboratory plasma experiment for studying magnetic dynamics of accretion discs and jets. *Mon. Not. R. Astron. Soc.* **334**, 257–261 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work supported by the US DOE, NSF and AFOSR.

Author Contributions A.L.M. performed the experiments and analysed data. A.L.M. and P.M.B. discussed and interpreted the results and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.L.M. (auna@caltech.edu).

Realization of three-qubit quantum error correction with superconducting circuits

M. D. Reed¹, L. DiCarlo², S. E. Nigg¹, L. Sun¹, L. Frunzio¹, S. M. Girvin¹ & R. J. Schoelkopf¹

Quantum computers could be used to solve certain problems exponentially faster than classical computers, but are challenging to build because of their increased susceptibility to errors. However, it is possible to detect and correct errors without destroying coherence, by using quantum error correcting codes¹. The simplest of these are three-quantum-bit (three-qubit) codes, which map a one-qubit state to an entangled three-qubit state; they can correct any single phase-flip or bit-flip error on one of the three qubits, depending on the code used². Here we demonstrate such phase- and bit-flip error correcting codes in a superconducting circuit. We encode a quantum state^{3,4}, induce errors on the qubits and decode the error syndrome—a quantum state indicating which error has occurred—by reversing the encoding process. This syndrome is then used as the input to a three-qubit gate that corrects the primary qubit if it was flipped. As the code can recover from a single error on any qubit, the fidelity of this process should decrease only quadratically with error probability. We implement the correcting three-qubit gate (known as a conditional-conditional NOT, or Toffoli, gate) in 63 nanoseconds, using an interaction with the third excited state of a single qubit. We find 85 ± 1 per cent fidelity to the expected classical action of this gate, and 78 ± 1 per cent fidelity to the ideal quantum process matrix. Using this gate, we perform a single pass of both quantum bit- and phase-flip error correction and demonstrate the predicted first-order insensitivity to errors. Concatenation of these two codes in a nine-qubit device would correct arbitrary single-qubit errors. In combination with recent advances in superconducting qubit coherence times^{5,6}, this could lead to scalable quantum technology.

Quantum error correction relies on detecting the presence of errors without gaining knowledge of the encoded quantum state. In the three-qubit error-correcting code, the subspace of the two additional ‘ancilla’ qubits uniquely encodes which of the four possible single-qubit errors has occurred, including the possibility of no flip. Crucially, errors consisting of finite rotations can also be corrected using these schemes because the error syndromes are allowed to be in superpositions of the possible outcomes, flipped and not flipped². Previous works implementing error correcting codes in liquid^{7–9} and solid-state¹⁰ NMR and with trapped ions^{11,12} have demonstrated two possible strategies for using the error syndromes. The first is to measure the ancillas (thereby projecting the syndrome) and use a classical logic operation to correct the detected error. This ‘feed-forward’ capability is challenging in superconducting circuits as it requires a fast and high-fidelity quantum non-demolition measurement, but is probably a necessary component to achieve scalable fault tolerance^{2,13}. The second strategy, as recently demonstrated with trapped ions¹² and used here, is to replace the classical logic with a quantum controlled-controlled NOT (CCNOT) gate that performs the correction coherently, leaving the entropy associated with the error in the ancilla qubits, which can then be reset¹⁴ if the code is to be repeated. The CCNOT gate performs exactly the action that would follow the measurement in the first scheme: flipping the primary qubit if and only if the ancillas encode the associated error syndrome.

The CCNOT gate is also vital for a wide variety of applications such as Shor’s factoring algorithm¹⁵ and has attracted substantial experimental interest, with recent implementations in linear optics¹⁶, trapped ions¹⁷ and superconducting circuits^{18,19}. Here we use the circuit quantum electrodynamics architecture²⁰ to couple four superconducting transmon qubits²¹ to a single microwave cavity bus²², where each qubit transition frequency can be controlled on nanosecond timescales with individual flux bias lines²³ and collectively measured by interrogating transmission through the cavity²⁴. (The details of the device can be found in Methods Summary and in ref. 3.) The frequencies of the qubits, labelled Q₁–Q₄, are tuned respectively to 6, 7, 7.85 and ~ 13 GHz, with Q₄ unused. In this Letter, we first demonstrate the three-qubit interaction used in the gate, which is an extension of interactions used in previous two-qubit gates^{3,23,25}, and show how this interaction can be used to create the desired CCNOT gate. We then verify its action and use it to demonstrate error correction for an error on a single qubit with the bit-flip code and then for simultaneous errors on all three qubits with the phase-flip code. We find a quadratic dependence of process fidelity on error probability, indicating that the algorithm is correcting errors as predicted.

Our three-qubit gate uses an interaction with the third excited state of one transmon. Specifically, it relies on the unique capability among computational states (eigenstates of the Pauli operator Z) of $|111\rangle$ to interact with the non-computational state $|003\rangle$ (the notation $|abc\rangle$ refers to the excitation levels of Q₁–Q₃, respectively). As the direct interaction of these states is prohibited to first order, we first transfer the quantum amplitude of $|111\rangle$ to the intermediate state $|102\rangle$, which itself couples strongly to $|003\rangle$. Calculated energy levels and time-domain data showing interaction between $|011\rangle$ and $|002\rangle$ (which is identical to that between $|111\rangle$ and $|102\rangle$ except for a 6-GHz offset) as a function of the flux bias on Q₂ are shown in Fig. 1a. Once the amplitude of $|111\rangle$ has been transferred to $|102\rangle$ with a sudden swap interaction, a three-qubit phase is acquired by moving Q₁ up in frequency adiabatically, near the avoided crossing with $|003\rangle$. Figure 1b shows the avoided crossing between these states as a function of the flux bias on Q₁. This crossing shifts the frequency of $|102\rangle$ relative to the sum of the frequencies of $|100\rangle$ and $|002\rangle$ to yield the three-qubit phase. The detailed procedure of the gate is shown in Fig. 2a, and is implemented in 63 ns. Further details can be found in Supplementary Information.

We demonstrate the gate by first measuring its classical action. The controlled-controlled phase (CCPhase) gate, which maps $|111\rangle$ to $-|111\rangle$, has no effect on pure computational states so we implement a CCNOT gate by concatenating pre- and post-gate rotations on Q₂, as indicated in the unshaded regions of Fig. 2a. Such a gate ideally swaps $|101\rangle$ and $|111\rangle$ and does nothing to the remaining states. To verify this, we prepare the eight computational states, implement the gate and measure its output using three-qubit state tomography³ to generate the classical truth table. The intended state is reached with $85 \pm 1\%$ fidelity on average. This measurement is sensitive only to classical action, however, so we complete our verification by performing full quantum process tomography on the CCPhase gate, which can detect

¹Departments of Physics and Applied Physics, Yale University, New Haven, Connecticut 06520, USA. ²Kavli Institute of Nanoscience, Delft University of Technology, 2628 CJ Delft, The Netherlands.

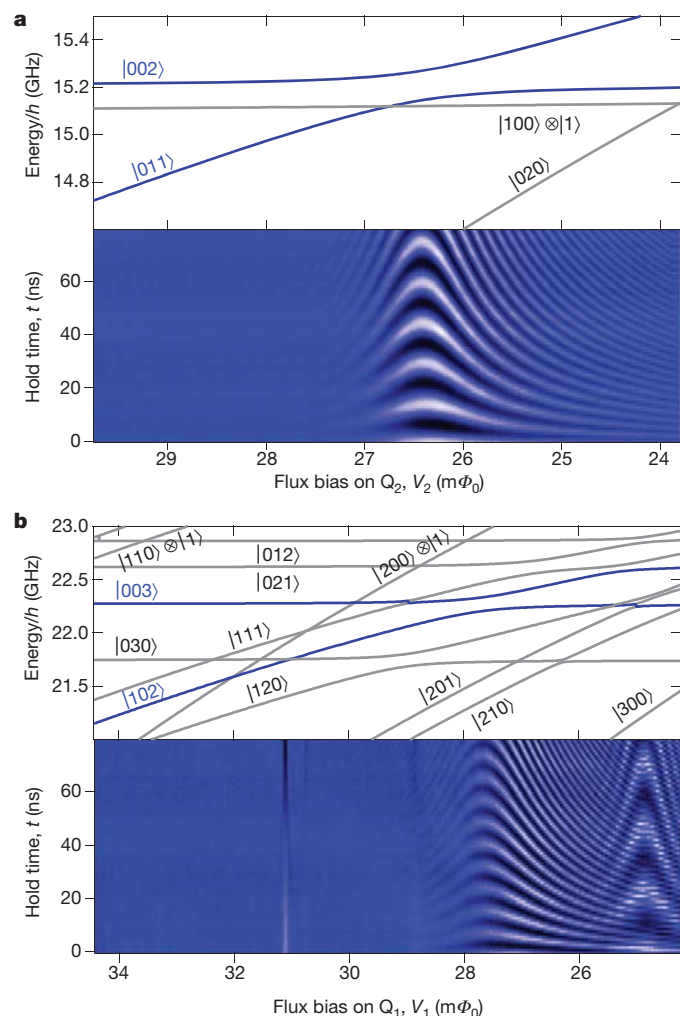


Figure 1 | Calculated energy spectra and time-domain measurements of the interactions used in the three-qubit gate. **a**, The energy spectrum of doubly excited states demonstrating the avoided crossing between $|011\rangle$ and $|002\rangle$ (identical to that between $|111\rangle$ and $|102\rangle$ except for a 6-GHz offset) is shown with both a numerical diagonalization of the system Hamiltonian (top) and a time-domain measurement as a function of the applied magnetic flux on Q_2 (bottom). Top: the frequencies for the involved eigenstates are blue and the non-interacting eigenstates of similar energy are grey. The notation $|abc\rangle \otimes |d\rangle$ indicates the excitation level of each qubit and the cavity photon number, respectively. When the second ket is omitted, $d = 0$. Bottom: the state $|011\rangle$ is prepared and a square flux pulse of duration t and amplitude V_2 is applied. Coherent oscillations produce a ‘chevron’ pattern, with darker colours corresponding to population left in $|002\rangle$. h , Planck’s constant. **b**, The spectrum of triply excited states showing the avoided crossing between $|102\rangle$ and $|003\rangle$ as a function of the flux bias on Q_1 is characterized in the same way as above. The state $|102\rangle$ is prepared by first making $|111\rangle$ and then performing the swap as described in Fig. 2. Many additional eigenstates are close in energy but are irrelevant because they do not interact with the populated states. The large avoided crossing between the relevant eigenstates that is used to produce an adiabatic three-qubit interaction happens near $28\text{ m}\Phi_0$ (where Φ_0 is the magnetic flux quantum). Extra lines near $31\text{ m}\Phi_0$ and $29\text{ m}\Phi_0$ are due to higher-order interactions predicted by the Hamiltonian ($|102\rangle$ with $|030\rangle$ and $|003\rangle$ with $|111\rangle$), as is the larger first-order interaction at $25\text{ m}\Phi_0$ ($|102\rangle$ with a hybridization of $|021\rangle$ and $|111\rangle$), but their effect on the protocol in Fig. 2 is negligible.

the evolution of quantum superpositions of computational states. This is done by preparing 64 input states that span the computational Hilbert space and by performing state tomography on the result of the gate’s action on each state. As detailed in Supplementary Information, we find a fidelity of $78 \pm 1\%$ to a process in which the spurious two-qubit phase between Q_1 and Q_3 is set to the independently

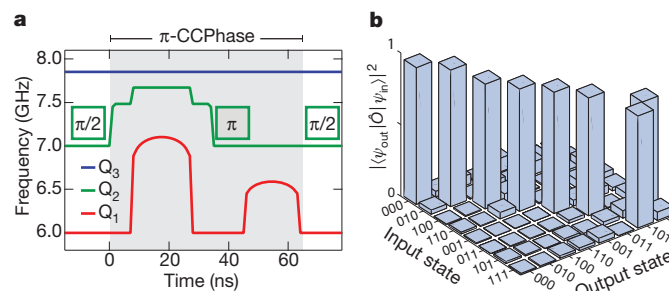


Figure 2 | Pulse sequence and classical action of the three-qubit gate. **a**, The frequencies of the three qubits and the locations of applied rotations during the three-qubit gate as functions of time. Shaded region: to produce the CCPhase interaction, Q_2 is first moved suddenly into resonance with the avoided crossing shown in Fig. 1a, which coherently transfers the population of $|111\rangle$ to $|102\rangle$ (and also that of $|011\rangle$ to $|002\rangle$) in 7 ns. Fine adjustments in the first point of the pulse compensate for finite pulse rise time and temporal precision. The frequency of Q_2 is then abruptly increased to where its two-qubit phase with Q_3 is cancelled during the gate by accumulating a multiple of 2π . The frequency of Q_1 is then increased adiabatically to initiate the interaction between $|102\rangle$ and $|003\rangle$. The duration and amplitude of this excursion is tuned to acquire a three-qubit phase of π . The population in $|102\rangle$ is then transferred back to $|111\rangle$ by reversing the swap procedure. Finally, the two-qubit phase between Q_1 and Q_2 is cancelled with an additional adiabatic interaction, which is sped up with a π -pulse on Q_2 at 37 ns (all rotations here are done about the x axis). The two-qubit phase between Q_1 and Q_3 is uncontrolled and there is an overall π -rotation of Q_2 , making this a π -CC- $e^{i\phi}$ Z gate, taking a total of 63 ns. Unshaded region: pre- and post-gate rotations on Q_2 appended to the CCPhase gate turn its action into that of a CCNOT gate, as described in Supplementary Information. **b**, The classical action of the CCNOT gate is measured by preparing the eight computational basis states, $|\psi_{in}\rangle$, and performing state tomography on the resulting state, $|\psi_{out}\rangle$, after applying the gate, \hat{O} , to them. The projection of these measurements to the computational basis states is taken to generate the displayed truth table. The fidelity to the expected action, where only the states $|101\rangle$ and $|111\rangle$ are swapped, is $85 \pm 1\%$. Full quantum process tomography of the gate is shown in Supplementary Information.

measured value of 57° (see Supplementary Information for an explanation of why this phase is irrelevant here). Owing to this extraneous phase, ϕ , the gate is most accurately described as a CC- $e^{i\phi}$ Z gate. The loss of fidelity is consistent with the expected energy relaxation of the three qubits during the 85-ns tomography procedure, which includes preparation and analysis pulses in addition to the gate, with some remaining error due to qubit transition frequency drift during the 90 min it takes to collect the full data set.

With our CCPhase gate in hand, we now demonstrate three-qubit error correction. We first examine the bit-flip code, which, as shown in Fig. 3a, starts by encoding the quantum state to be protected in a three-qubit entangled state² through the use of conditional phase (CPhase) gates. The state $\alpha|0\rangle + \beta|1\rangle$ is encoded as $\alpha|000\rangle + \beta|111\rangle$, which has the property that the value of any two-qubit ZZ product is $+1$ regardless of the values of α and β . (For quantum states on the equator of the Bloch sphere, $|\alpha| = |\beta| = 1/\sqrt{2}$ and the encoding is a maximally entangled three-qubit Greenberger–Horne–Zeilinger state^{3,4,26} that we independently measure to have a state fidelity of $89 \pm 1\%$.) If any single qubit is flipped, one or more of the ZZ products will flip sign as well. For example, if Q_1 were flipped, the Z_1Z_2 product would become -1 whereas the Z_2Z_3 product would remain $+1$, uniquely indicating that Q_1 needs to be corrected. Indeed, the four possible combinations of Z_1Z_2 and Z_2Z_3 exactly encode the possible single bit flips, including the possibility of no flip. In a fault-tolerant code, these products would be stored in separate qubits for later measurement², but here we instead reverse the encoding so that the ancillas Q_1 and Q_3 can no longer witness bit-flip errors and instead store the values of the two ZZ products. These ancillas are then used as the control bits for the CCNOT gate described above, so that Q_2 will be flipped back if and only if both ancillas are excited, which indicates that Q_2 was flipped. The detailed

evolution of the qubits during the error correction procedure can be found in Supplementary Information.

Whereas errors on classical bits are discrete, quantum error correction must be able to correct arbitrary rotations as well as complete flips because superpositions of states are allowed. Remarkably, the code described above already satisfies this criterion. If an error causes a rotation θ on Q_2 , the quantum state after decoding will be $\sqrt{1-p}(\alpha|0\rangle + \beta|1\rangle) \otimes |00\rangle + \sqrt{p}(\beta|0\rangle + \alpha|1\rangle) \otimes |11\rangle$, where $p = \sin^2(\theta/2)$ is the effective probability of a full flip and where we have listed first the state of Q_2 followed by those of Q_1 and Q_3 for notational simplicity. That is, the state will be a superposition of Q_2 in the correct state with the ancillas indicating no error plus Q_2 flipped with the ancillas indicating as such. The application of the CCNOT gate to this state will successfully correct it because it acts only on the subspace where both ancillas are excited, making the state $(\alpha|0\rangle + \beta|1\rangle) \otimes (\sqrt{1-p}|00\rangle + \sqrt{p}|11\rangle)$. (Alternatively, if the ancilla

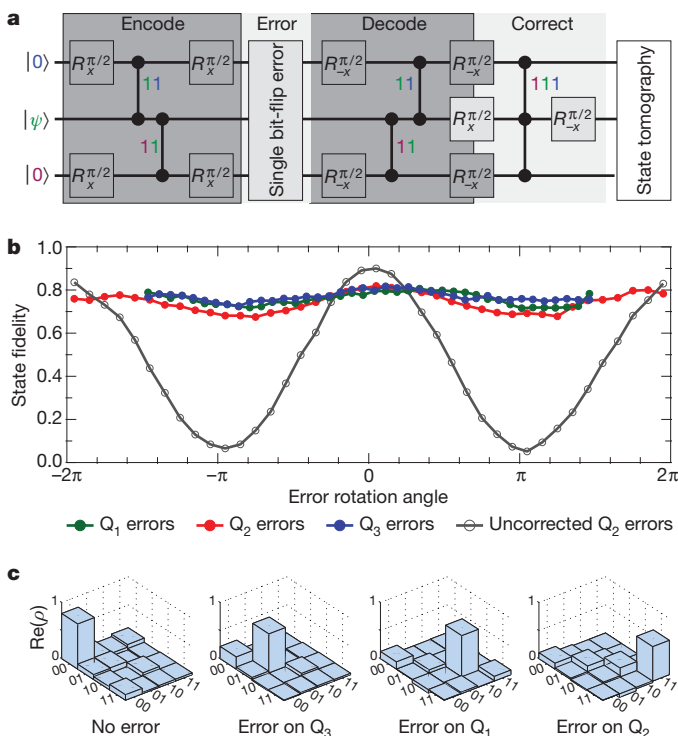


Figure 3 | Bit-flip error correction demonstrating recovery from a single arbitrary rotation. **a**, The error correction protocol starts by encoding the quantum state to be protected in a three-qubit state by entangling the two ancilla qubits, Q_1 and Q_3 , with Q_2 through the use of two CPhase gates (vertical lines terminating in solid circles) and $\pi/2$ -rotations ($R_{\pi/2}$ is a single-qubit rotation, where π indicates the rotation axis and α is the rotation angle). The number adjacent to each CPhase gate indicates which state receives a phase shift of π . A single y -rotation error of a known angle is then performed on a single qubit (as explained in Supplementary Information, this is compiled together with other rotations when acting on the ancillas). The state is then decoded, leaving the ancillas in a product state indicating which single-qubit error occurred. For finite rotations, the ancillas will be in a superposition of states in which the error did and, respectively, did not occur, with each tensor multiplied with the associated single-qubit state of Q_2 . If an error occurred on Q_2 , the CCNOT gate implemented with our CCPhase gate (represented by three solid circles linked by a vertical line) at the end of the code will correct it. We then perform three-qubit state tomography to verify the result. **b**, State fidelity to the created state $|\psi\rangle = |+\rangle$ after causing an error on one of the qubits, with and without error correction. Ideally, the error-corrected curves would be horizontal lines at unit fidelity. Finite excited-state lifetimes cause oscillations and lower fidelity because errors change the excitation level of the system. **c**, Two-qubit density matrices (ρ) of the ancillas after each of the four possible full bit-flip errors has occurred. The fidelities of these states to the ideal error syndromes, $|00\rangle$, $|01\rangle$, $|10\rangle$ and $|11\rangle$, are respectively 81.3%, 69.7%, 73.1% and 61.2%.

qubits were measured, they would project the state onto one of those possibilities, essentially forcing the computer to ‘decide’ whether a full flip had occurred.) We demonstrate this with the procedure shown in Fig. 3a for the state $|\psi\rangle = |+\rangle$ (the positive eigenstate of the Pauli operator X), performing single deterministic rotations of a known angle on each of the three qubits to simulate errors. As shown in Fig. 3b, we compare this with the case of uncorrected errors on Q_2 . Ideally, the error-corrected curves would have unit fidelity and be independent of θ , but they are slightly lower in fidelity and oscillate in θ owing to finite coherence. They are, however, substantially improved relative to the uncorrected case, demonstrating that the errors are in fact being ameliorated. As shown in Fig. 3c, we also measure the two-qubit density matrix of the ancilla qubits after each of the four possible full bit-flip errors, showing that they end up in a computational product state correctly indicating the induced error.

In real physical systems, errors occur at approximately the same rate on all constituent qubits rather than on one at a time. The correction

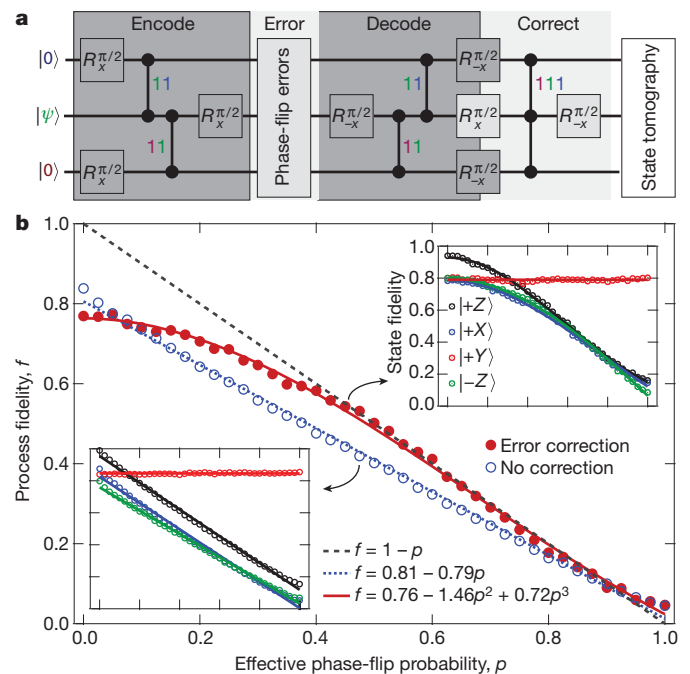


Figure 4 | Demonstration of first-order insensitivity to simultaneous phase-flip errors. **a**, The phase-flip error correction protocol differs from the bit-flip protocol described in Fig. 3a only by single-qubit gates. Those gates effectively rotate the coordinate system, mapping phase flips to bit flips, and vice versa, so the remainder of the procedure is exactly the same as in the bit-flip case². We perform errors on all three qubits simultaneously with z gates of known rotation angle, which is equivalent to phase-flip errors with probability $p = \sin^2(\theta/2)$. As with the bit-flip code, if a single error has occurred on the primary qubit, the CCNOT gate at the end of the code will undo it. **b**, Fidelity of the process matrix of the protected qubit to the identity operation plotted as a function of p . As the code corrects only single errors, it will fail on the three-qubit subspace where more than one has occurred, which happens with a probability $3p^2 - 2p^3$. The coefficients here are reduced for processes with finite fidelity. The process fidelity is fitted with $f = (0.760 \pm 0.005) - (1.46 \pm 0.03)p^2 + (0.72 \pm 0.03)p^3$. If a linear term is allowed, its best-fit coefficient is 0.03 ± 0.06 . We compare this with the case of no error correction to simulate the improvement made when the decoherence of Q_2 is normalized away (blue symbols). We also plot the simulated fidelity of a perfect but non-corrected system (black dashed line), which indicates that for our gate fidelities we do not show a net improvement for artificial errors. Insets: the constituent state fidelities of the four basis states used to produce the process fidelity data in the case with error correction (right) and in the case with no correction (left). The x axes of the plots are the same as the main panel, and they share the same y axis. The state $|+\rangle$ (the positive eigenstate of the Pauli operator Y) is immune to errors because its encoded state is an eigenstate of single, double and triple qubit phase flips.

scheme will only succeed, therefore, on the three-qubit subspace with zero or one errors. The probability of more than one error occurring is $3p^2 - 2p^3$, where p is the single-qubit error rate², so the fidelity of error correction should be $1 - 3p^2 + 2p^3$. For a scheme with gate fidelity limited by decoherence, the coefficients of the quadratic and cubic terms will be smaller but, crucially, any linear dependence on p will be strongly suppressed. If the error rates for each qubit were different, these coefficients would again be modified but any linear dependence would still be abated. For the sake of completeness, here we use the phase-flip code, which differs from the previously discussed bit-flip code by only single-qubit rotations, as shown in Fig. 4a. This difference can be viewed as a rotation of the coordinate system, converting phase flips to bit flips and vice versa, so the remainder of the code is exactly the same as the previous case^{2,12,27}. Phase errors of known rotation angle are applied by rotating the frame of reference of subsequent x and y rotations. As shown in Fig. 4b, we measure the process fidelity of the error correction scheme as a function of p and compare this with the case of no error correction in which identical single-qubit rotations are applied to Q_2 but the ancillas are not involved (this comparison is without gates, but with appropriate delays to have the same total procedure duration, to indicate the lack of fidelity due to the decoherence of Q_2). Whereas without error correction we find a purely linear dependence on p , with the correction applied the data are extremely well modelled by only quadratic and cubic terms, demonstrating the desired first-order insensitivity to errors. We have therefore realized a successful implementation of quantum error correction, although improvement of the fidelity of a real physical process will require considerable advances in both gate fidelity and device complexity.

We have realized both bit- and phase-flip error correction in a superconducting circuit. In doing so, we have tested both main conceptual components of the nine-qubit Shor code¹, which can defend against arbitrary single-qubit errors by concatenating the bit- and phase-flip codes. The implementation relies on our efficient three-qubit gate, which uses non-computational states in the third excitation manifold of our system, demonstrating that the simple Hamiltonian of the system accurately predicts the dynamics even at these high excitation levels. The gate takes approximately half the time of an equivalent construction with one- and two-qubit gates. We expect it to work between any three nearest-neighbour qubits in frequency regardless of the number of qubits sharing the bus, as interactions involving other qubits will be first-order prohibited.

METHODS

Arbitrary qubit rotations around the x and y axes of the Bloch sphere are performed with pulse-shaped resonant microwave tones. Rotations around the z axis are made by rotating the reference phase of subsequent x and y pulses. One-qubit dynamical phases resulting from flux excursions are measured with modified Ramsey experiments comparing the phase acquired by an unmodified prepared state with the phase acquired by that same state after a flux pulse, and are cancelled with z rotations. Two- and three-qubit phases are measured with a similar Ramsey experiment comparing the phase acquired when a control qubit is in its ground state with the phase acquired when it is in an excited state. For example, the two-qubit phase between Q_2 and Q_3 is measured by preparing Q_3 along the y axis and Q_2 either in its ground or excited state and then performing the flux pulse in both cases. The single-qubit phase of Q_3 is the same for both states, so the two-qubit phase is directly measurable as the difference in phase between them. All phases are initially tuned to within 1° , limited by the resolution of control equipment and drifts of system parameters such as the qubit transition frequencies.

Received 21 September; accepted 7 December 2011.

Published online 1 February 2012.

- Shor, P. W. Scheme for reducing decoherence in quantum computer memory. *Phys. Rev. A* **52**, R2493–R2496 (1995).

- Nielsen, M. A. & Chuang, I. L. *Quantum Computation and Quantum Information* (Cambridge Ser. Inf. Nat. Sci, Cambridge Univ. Press, 2000).
- DiCarlo, L. *et al.* Preparation and measurement of three-qubit entanglement in a superconducting circuit. *Nature* **467**, 574–578 (2010).
- Neeley, M. *et al.* Generation of three-qubit entangled states using superconducting phase qubits. *Nature* **467**, 570–573 (2010).
- Paik, H. *et al.* Observation of high coherence in Josephson junction qubits measured in a three-dimensional circuit QED architecture. *Phys. Rev. Lett.* **107**, 240501 (2011).
- Kim, Z. *et al.* Decoupling a Cooper-pair box to enhance the lifetime to 0.2 ms. *Phys. Rev. Lett.* **106**, 120501 (2011).
- Cory, D. *et al.* Experimental quantum error correction. *Phys. Rev. Lett.* **81**, 2152–2155 (1998).
- Knill, E., Laflamme, R., Martinez, R. & Negrevergne, C. Benchmarking quantum computers: the five-qubit error correcting code. *Phys. Rev. Lett.* **86**, 5811–5814 (2001).
- Boulant, N., Viola, L., Fortunato, E. & Cory, D. Experimental implementation of a concatenated quantum error-correcting code. *Phys. Rev. Lett.* **94**, 130501 (2005).
- Moussa, O., Baugh, J., Ryan, C. A. & Laflamme, R. Demonstration of sufficient control for two rounds of quantum error correction in a solid state ensemble quantum information processor. *Phys. Rev. Lett.* **107**, 160501 (2011).
- Chiaverini, J. *et al.* Realization of quantum error correction. *Nature* **432**, 602–605 (2004).
- Schindler, P. *et al.* Experimental repetitive quantum error correction. *Science* **332**, 1059–1061 (2011).
- Shor, P. W. In *Proc. 37th Symp. Foundations Comput.* 56–65 (IEEE, 1996); preprint at (<http://arxiv.org/abs/quant-ph/9605011>) (1996).
- Reed, M. D. *et al.* Fast reset and suppressing spontaneous emission of a superconducting qubit. *Appl. Phys. Lett.* **96**, 203110 (2010).
- Shor, P. W. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM J. Sci. Stat. Comput.* **26**, 1484–1509 (1997).
- Lanyon, B. P. *et al.* Simplifying quantum logic using higher-dimensional Hilbert spaces. *Nature Phys.* **5**, 134–140 (2009).
- Monz, T. *et al.* Realization of the quantum Toffoli gate with trapped ions. *Phys. Rev. Lett.* **102**, 040501 (2009).
- Mariantoni, M. *et al.* Implementing the quantum von Neumann architecture with superconducting circuits. *Science* **334**, 61–65 (2011).
- Fedorov, A., Steffen, L., Baur, M. & Wallraff, A. Implementation of a Toffoli gate with superconducting circuits. *Nature* **481**, 170–172 (2012).
- Wallraff, A. *et al.* Strong coupling of a single photon to a superconducting qubit using circuit quantum electrodynamics. *Nature* **431**, 162–167 (2004).
- Schreier, J. A. *et al.* Suppressing charge noise decoherence in superconducting charge qubits. *Phys. Rev. B* **77**, 180502 (2008).
- Majer, J. *et al.* Coupling superconducting qubits via a cavity bus. *Nature* **449**, 443–447 (2007).
- DiCarlo, L. *et al.* Demonstration of two-qubit algorithms with a superconducting quantum processor. *Nature* **460**, 240–244 (2009).
- Reed, M. *et al.* High-fidelity readout in circuit quantum electrodynamics using the Jaynes-Cummings nonlinearity. *Phys. Rev. Lett.* **105**, 173601 (2010).
- Strauch, F. *et al.* Quantum logic gates for coupled superconducting phase qubits. *Phys. Rev. Lett.* **91**, 167005 (2003).
- Greenberger, D. M., Horne, M. A. & Zeilinger, A. in *Bell's Theorem, Quantum Theory and Conceptions of the Universe* (ed. Kafatos, M.) 69–72 (Kluwer Academic, 1989).
- Tomberg, L., Wallquist, M., Johansson, G., Shumeiko, V. & Wendin, G. Implementation of the three-qubit phase-flip error correction code with superconducting qubits. *Phys. Rev. B* **77**, 214528 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank G. Kirchmair, M. Mirrahimi, I. Chuang and M. Devoret for discussions. We acknowledge support from LPS/NSA under ARO contract no. W911NF-09-1-0514 and from the NSF under grants no. DMR-0653377 and no. DMR-1004406. Additional support was provided by CNR-Istituto di Cibernetica, Pozzuoli, Italy (L.F.), the Swiss NSF (S.E.N.) and the Dutch NWO (L.D.C.).

Author Contributions M.D.R. carried out measurements and performed data analysis. L.D.C. designed the three-qubit gate and conducted initial measurements. L.S. provided further experimental contributions. S.E.N. and S.M.G. provided theoretical support. L.F., L.D.C. and L.S. fabricated the devices. M.D.R. wrote the manuscript, with feedback from all authors. S.M.G. and R.J.S. designed and supervised the project.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.D.R. (matthew.reed@yale.edu) or R.J.S. (robert.schoelkopf@yale.edu).

Origin of Columbia River flood basalt controlled by propagating rupture of the Farallon slab

Lijun Liu¹ & Dave R. Stegman¹

The origin of the Steens–Columbia River (SCR) flood basalts, which is presumed to be the onset of Yellowstone volcanism, has remained controversial, with the proposed conceptual models involving either a mantle plume^{1–5} or back-arc processes^{6–8}. Recent tomographic inversions based on the USArray data reveal unprecedented detail of upper-mantle structures of the western USA⁹ and tightly constrain geodynamic models simulating Farallon subduction, which has been proposed to influence the Yellowstone volcanism^{5,6}. Here we show that the best-fitting geodynamic model¹⁰ depicts an episode of slab tearing about 17 million years ago under eastern Oregon, where an associated sub-slab asthenospheric upwelling thermally erodes the Farallon slab, leading to formation of a slab gap at shallow depth. Driven by a gradient of dynamic pressure, the tear ruptured quickly north and south and within about two million years covering a distance of around 900 kilometres along all of eastern Oregon and northern Nevada. This tear would be consistent with the occurrence of major volcanic dikes during the SCR–Northern Nevada Rift flood basalt event both in space and time. The model predicts a petrogenetic sequence for the flood basalt with sources of melt starting from the base of the slab, at first remelting oceanic lithosphere and then evolving upwards, ending with remelting of oceanic crust. Such a progression helps to reconcile the existing controversies on the interpretation of SCR geochemistry and the involvement of the putative Yellowstone plume. Our study suggests a new mechanism for the formation of large igneous provinces.

The SCR igneous province of the Pacific Northwest represents one of the largest continental flood basalt events, with a total eruption volume of around 230,000 km³ over approximately two million years (Myr) (ref. 3). This massive, fast eruption seems to favour a mantle plume origin^{1–3}, but a plume model cannot address why most SCR flood basalt erupted in a north–south-oriented region perpendicular to the subsequent Yellowstone hotspot track along the eastern Snake River plain (Fig. 1). Recent models trying to explain this complexity include spreading of the plume head along a lithospheric gradient⁴ and lateral deflection of the plume conduit due to mantle flow⁵. Other workers, however, dismissed the plume hypothesis as conjecture, and argued that the SCR event could have been a result of shallow-mantle processes, such as back-arc extension⁶ or small-scale convection⁷ or lithosphere delamination⁸. These conceptual models are all based on some aspects of surface geologic features, but the implied underlying mantle dynamics differ significantly from each other. A better understanding of the SCR flood basalt formation, therefore, requires an improved knowledge of mantle processes during the mid-Miocene epoch (about 16 Myr ago), especially given that Farallon subduction adequately explains both the observed surface plate kinematics and continental deformation within the western United States.

A promising way to infer past mantle dynamics is by predicting its present-day state through geodynamic modelling using a technique called data assimilation, which can be either sequential¹¹ or variational¹². Here we adopt the sequential technique and assimilate plate motion history, palaeo-seafloor ages and palaeo-geometry of plate boundaries into a single geodynamic model¹⁰. The model integrates

from 40 Myr ago to the present, through which we try to predict the observed mantle structures beneath the western USA outlined with increasing detail by a sequence of tomographic inversions based on data from the USArray⁹ (see Supplementary Fig. 1 for more tomography models). The most robust seismic feature, a 500-km-wide columnar fast anomaly extending from 300 to 600 km depth below Nevada and western Utah, was found to be a segmented piece of the Farallon slab that was folded upward along its edges by toroidal mantle flows during its descent (Supplementary Fig. 2). Tracking backward in time, the initial break-off of this slab happened around 17 Myr ago at shallow depth (Figs 1 and 2). This is because the shrinking width of the

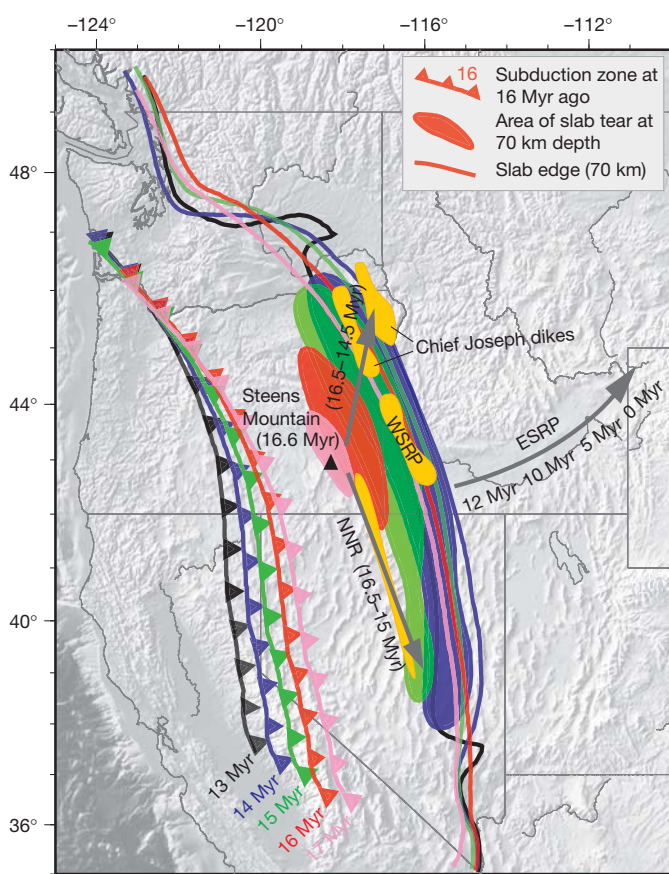


Figure 1 | Development of the Farallon slab rupture beneath the western USA during the mid-Miocene epoch. Geometry of Farallon subduction at different times (corresponding to different colours, ages in Myr as shown) is projected onto North America. Both the slab edge (solid lines) and slab gap (filled area) are at 70 km depth, outlined by an isotherm (-50°C relative to ambient mantle). Major SCR volcanic dikes following ref. 27 are shown as yellow patterns (with ages in Myr shown). Grey arrows indicate the direction of age (shown, in Myr) progression of surface eruptions. WSRP, western Snake River plain; ESRP, eastern Snake River plain; NNR, northern Nevada rift zone.

¹IGPP, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California 92093, USA.

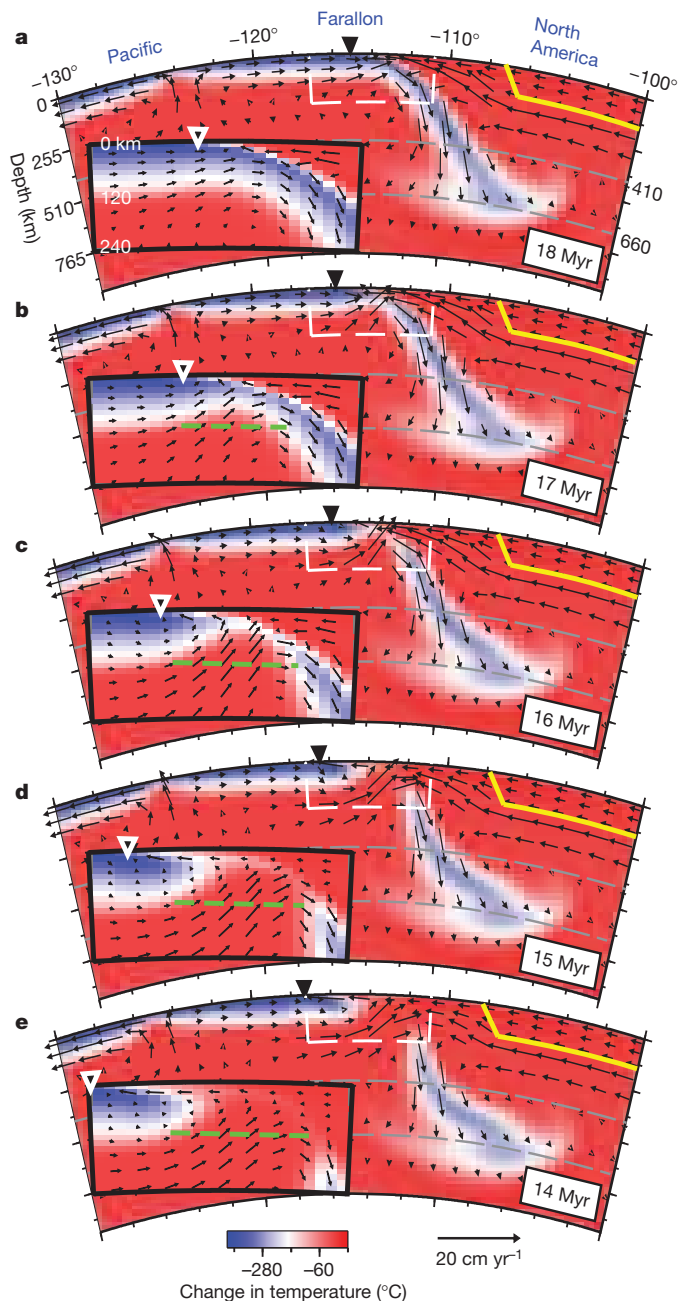


Figure 2 | Mantle flow associated with the slab tearing at latitude 44° N from 18 to 14 Myr ago. Axes in b–e are the same as those in a. Solid yellow outlines represent the high-viscosity North America continent. Inverse black triangles mark trench locations with time. Insets are zoom-in plots of the subduction zones outlined by the white dashed boxes. The green dashed line represents the depth of 110 km, above which melting would occur¹⁸. We do not explicitly model melt generation because parameterizations of melting processes commonly implemented are applicable to melting of ambient mantle material and are not relevant to different compositions such as oceanic lithosphere and oceanic crust, as in the scenario presented here. Black arrows show mantle velocities, and the change in temperature is relative to the ambient mantle.

oceanic plate¹³ gradually built up dynamic pressure beneath the middle part of the subducting slab, where the increased pressure caused the slab to flatten¹⁰ (see Fig. 2a for example). Consequently, the flattening slab slowed down the forward plate motion, and sped up back-arc extension, as predicted by fully dynamic subduction models¹⁴. The reduction in plate motion and regionally focused sub-slab dynamic pressure generated an extension in the central part of the slab hinge, which, combined with thermal erosion due to sub-slab upwelling,

caused the weak slab to stretch and eventually break in the middle¹⁰ (Figs 2 and 3), similar in some ways to slab tear migrations proposed for Mediterranean slabs¹⁵.

To understand the possible surface effects of this slab tear, we project the Farallon subduction system onto the reference frame of North America, neglecting the internal deformation of the western USA due to Basin and Range extension¹⁶. The initial slab tear occurred as early as about 17 Myr ago, which, in a map view, formed a trench-parallel slab gap (defined by an isotherm at a depth of 70 km) inside the back-arc region beneath southeastern Oregon (Fig. 1). This gap coincides with the location of Steens Mountain, which recorded the initial phase of eruption during the SCR flood basalt event. The slab tear ruptured quickly to north and south along the trench-parallel direction (Fig. 3), while at the same time the progressive flattening shifted the slab gap to the east (Figs 1 and 2). By 15 Myr ago, the gap occupied all of east Oregon, southwest Idaho and north Nevada (Fig. 1). This propagating pattern of slab tear correlates with the sequence of major mid-Miocene volcanic dikes both in space and time. From south to north with decreasing age, the western Snake River plain and Chief Joseph dike swarms hosted an increasing amount of magma outpouring, forming the Imnaha and Grande Ronde magmatic provinces⁴, consistent with the northward rupturing and widening of the slab gap. Meanwhile, the southward slab rupturing with a narrower gap also explains the rapid formation (within about 1.5 Myr) of the northern Nevada rift zone²⁷ and the observed limited magma eruption along this region. Furthermore, the reduction of trenchward wedge flow inside Oregon after 17 Myr ago (Figs 2 and 3) provides a physical explanation for the reduced magmatic activities along the Oregon coastal arc since this time¹⁷. About 14 Myr ago, the Farallon slab was largely separated into two pieces along the down-dip direction, the landward extent of flat slab underplating started to retreat to the west (Figs 1 and 2), and sub-slab pressure was largely equilibrated with the mantle wedge, consistent with termination of the SCR eruption around this time⁴.

Another important condition for making flood basalts is sufficient subsurface mantle upwelling, which is a salient feature of the mantle plume model. We also observe a strong focused mantle upwelling below the slab hinge, driven by the excess dynamic pressure beneath the subducting plate (Figs 2 and 3). Melting in this setting can reasonably be expected, because melts are generated at depths shallower than 110 km along forced upwellings in oceanic environments¹⁸. This fast upwelling flow advectively erodes the slab isotherm, leading to a progressively shallower asthenosphere, which suggests that oceanic lithosphere would melt first, followed by melting of oceanic crust. The modelled slab initially acted as both a thermal and mechanical barrier to upward flow and inhibited melting, as can be seen from 18 Myr ago, when the upwelling simply flattened the slab (Fig. 2a, inset). Subsequent thermal erosion of the oceanic lithosphere (Fig. 2b) would allow melt production under the Steens Mountain and Imnaha provinces (Table 1). By 16 Myr ago, continued thermal erosion melted away the entire oceanic lithosphere (Fig. 2c), where the fast melting of oceanic crust encountering the high mantle temperature should have triggered the massive magma outpouring of Grande Ronde⁴. The sustained cooling within the upwelling due to progressive melting seems to explain the marked decrease in magma volume during the formation of the Imnaha province (Table 1) and the step-function change in chemistry at the Imnaha–Grande Ronde boundary¹⁹. The upward corner flow inside the mantle wedge may also have facilitated melt generation, but its volume contribution is probably minor, given the much colder environment below the overriding plate, an aspect not explicitly represented in our convection model (Figs 2 and 3).

The geochemical data, essential for determining the source rocks, have been used both for and against a plume signature in the SCR lavas^{1,3,19–21}. Although origins of both major and trace elements are still widely debated among existing conceptual models, there are certain points of consensus (Table 1). (1) Steens Mountain lavas are derived from an ultra-depleted mantle source²⁰, probably “owing to repetitive

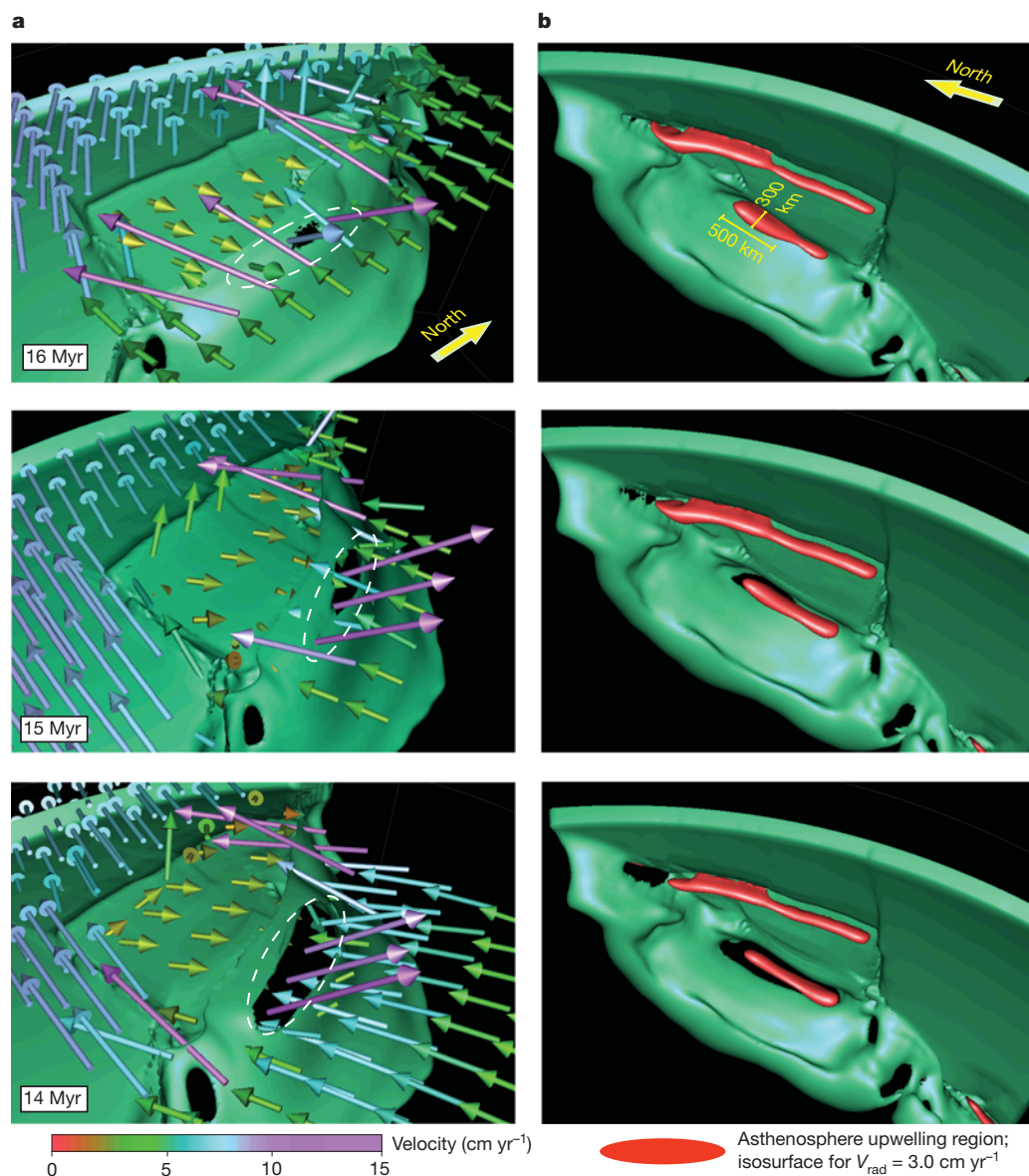


Figure 3 | Three-dimensional view of the tearing slab. **a**, Top view of an isothermal slab surface (-100°C relative to ambient temperature) and the velocity field. The velocity vectors are on the same spherical plane across all surface plates. The white dashed oval outlines a region where asthenosphere upwelling occurred. We note that the upwellings are restricted to below the

extraction of trace-element magmas²¹. (2) The Grande Ronde lavas have the highest silica content of all SCR lavas¹⁹, and are best explained by melting of a mafic crust²². (3) The SCR formations experienced a progressive assimilation of material from the overriding continent and from subducted sediments^{1,19,20}.

Farallon slab where velocity vectors originate (see Fig. 2 for detailed flow fields). **b**, Bottom view showing distribution of asthenosphere upwellings beneath the mid-ocean ridge (long red shape) and propagating slab tear (short red shape). The region inside the red surface represents mantle upwelling with radial velocity magnitudes $>3\text{ cm yr}^{-1}$.

These three features are consistent with our proposed bottom-up melting model as follows. The earliest eruption at Steens Mountain is derived from the depleted subducting oceanic lithosphere, whose melts also tend to be low in silica content. The upward melting across the subducting slab will eventually melt its mafic crust in an avalanche

Table 1 | Summary of SCR basalts and proposed mechanisms

SCR formation ⁴	Geochemical property ^{1,19–21}	Volume ^{3,4} (km ³)	Source composition	Unreconciled aspects of previous models
Steens (16.6–16.2 Myr)	Low silica content; high ϵ_{Nd} , low $^{87}\text{Sr}/^{86}\text{Sr}$, incompatible element depletion	60,000	Subducting oceanic lithosphere	Highly depleted magma source (plume-head models ^{1–5} , lithosphere-plume ¹⁹ or slab-plume ²⁵ interactions)
Imnaha (16.3–16.1 Myr)	Excess $^{206,208}\text{Pb}/^{204}\text{Pb}$, excess Th, Nb and $^3\text{He}/^4\text{He}$	10,000	Subducting oceanic lithosphere and sediments or mantle plume	
Grande Ronde (16.1–15.0 Myr)	Silica saturated; low ϵ_{Nd} , high $^{87}\text{Sr}/^{86}\text{Sr}$, chemically homogeneous, incompatible element enrichment	150,000	Subducting oceanic crust and sediments and Archean mantle lithosphere	High SiO_2 and homogeneity (plume-head models ^{1–5} , lithosphere-plume ¹⁹ or slab-plume ²⁵ interactions) and large volume and high SiO_2 (back-arc processes ^{6–8})

fashion, which best explains Grande Ronde's homogeneous composition, saturation in silica content, and high-degree of incompatible elements enrichment²⁰. With the mantle isotherm shifted towards shallower depths, a progressive involvement of sediments from the subducted sea floor and of the overriding lithosphere and crust is expected, which explains the transitional composition of Imnaha lavas^{19,20}. Our proposed mechanism, however, does not preclude the existence of secondary crustal processes, such as crystal fractionation, before surface eruption.

Of all SCR formations, elevated ³He/⁴He ratios in a subpopulation of Imnaha basalts are the only ones that might indicate a mantle plume signature²¹. However, involvement of a plume head should lead to both highly heterogeneous and enriched compositions throughout SCR lavas, which is contrary to observations and suggests that if a mantle plume ever existed, it played a very minor part in defining the SCR flood basalt. The remarkable spatial/temporal correlations shown in Fig. 1 and the consistency with geochemistry (Table 1) strongly favour a slab-rupture model over those with a plume-head origin. Furthermore, Palaeocene–Eocene-aged coastal ranges are interpreted as an accreted island chain formed by the Yellowstone hotspot^{23,24} and provide possible evidence for a long-lived Yellowstone plume conduit that pre-dates the SCR event. We speculate that the Farallon slab captured the plume conduit and caused a short hiatus of Yellowstone-related volcanism during the early Miocene epoch, consistent with the plume-affinity of Imnaha basalts resulting from a northeast-deflected plume conduit²⁵. Subsequently, accompanying the slab rupture, the plume conduit was reestablished to its original position along the ESRP^{2,4}, where a slab gap has existed throughout the upper mantle till the present day¹⁰, and provided a path for the plume conduit to reach the surface (Supplementary Fig. 2). However, the interactions of this putative plume conduit with the Farallon slabs remain to be clarified.

METHODS SUMMARY

To explain the complex present-day mantle structure beneath the western USA, we use forward geodynamic models to simulate Farallon subduction during the past 40 Myr (ref. 10). The models are consistent with several types of palaeo-records, where both the initial and boundary conditions are observationally constrained. The thermal structure of the oceanic plate is according to palaeo-sea floor ages. With a high numerical resolution of down to 7 km, the models can effectively represent sophisticated thermal and rheological structures along plate boundaries, including mid-ocean ridges, transform faults, and subduction zones. The viscosity profiles of these fine features, especially those along convergent plate boundaries, are constrained such that the subducting slab hinge closely follows the time history of observed trench locations. We also introduce a pseudo free-surface function such that subduction occurs more naturally than in models without this function.

The Mid-Miocene slab tear event as described in this paper was critical for the formation of the present-day mantle structure. We find that although the exact shape and position of present mantle structure are very sensitive to the magnitude of mantle and slab viscosities¹⁰, the slab tear starting at around 17 Myr ago was not: the tear always occurred with similar timing and location even when mantle viscosities varied by one order of magnitude around the best-fit values¹⁰. This slab tear, however, no longer occurred when we excluded the pseudo free-surface function in the calculation, indicating that the near-trench sub-slab dynamic pressure controls formation of the slab tear (Supplementary Fig. 3). Of all the models, the one that best predicts the seismic image also seems to match the SCR flood basalt stratigraphy best, which provides another validation for the geodynamic model.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 5 July; accepted 2 December 2011.

- Brandon, A. D. & Goles, G. G. A Miocene subcontinental plume in the Pacific Northwest: geochemical evidence. *Earth Planet. Sci. Lett.* **88**, 273–283 (1988).

- Pierce, K. L. & Morgan, L. A. in *Regional Geology of Eastern Idaho and Western Wyoming* (eds Link, P. K. et al.) **179**, 1–53 (Geological Society of America Memoir, 1992).
- Hooper, P. R., Camp, V. E., Reidel, S. P. & Ross, M. E. The origin of the Columbia River flood basalt province: plume versus nonplume models. *GSA Spec. Pap.* **430**, 635–668 (2007).
- Camp, V. E. & Ross, M. E. Mantle dynamics and genesis of mafic magmatism in the intermontane Pacific Northwest. *J. Geophys. Res.* **109**, B08204 (2004).
- Smith, R. B. et al. Geodynamics of the Yellowstone hotspot and mantle plume: seismic and GPS imaging, kinematics, and mantle flow. *J. Volcanol. Geotherm. Res.* **188**, 26–56 (2009).
- Carlson, R. W. & Hart, W. K. Crustal genesis on the Oregon plateau. *J. Geophys. Res.* **92**, 6191–6206 (1987).
- Christiansen, R. L., Foulger, G. R. & Evans, J. R. Upper mantle origin of the Yellowstone hot spot. *Geol. Soc. Am. Bull.* **114**, 1245–1256 (2002).
- Hales, T. C., Abt, D. L., Humphreys, E. D. & Roering, J. J. Delamination origin for the Columbia River flood basalts and Wallowa Mountain uplift in NE Oregon, USA. *Nature* **438**, 842–845 (2005).
- Sigloch, K. Mantle provinces under North America from multifrequency P wave tomography. *Geochem. Geophys. Geosyst.* **12**, Q02W08 (2011).
- Liu, L. & Stegman, D. R. Segmentation of the Farallon slab. *Earth Planet. Sci. Lett.* **311**, 1–10 (2011).
- Bunge, H.-P. & Grand, S. P. Mesozoic plate-motion history below the northeast Pacific Ocean from seismic images of the subducted Farallon slab. *Nature* **405**, 337–340 (2000).
- Liu, L., Spasojević, S. & Gurnis, M. Reconstructing Farallon plate subduction beneath North America back to the Late Cretaceous. *Science* **322**, 934–938 (2008).
- Atwater, T. & Stock, J. in *Integrated Earth and Environmental Evolution of the Southwestern United States* (eds Ernst, W. G. & Nelson, C. A.) 393–420 (Bellwether Publishing, 1998).
- Schellart, W. P., Stegman, D. R., Farrington, R. J., Freeman, J. & Moresi, L. Cenozoic tectonics of western North America controlled by evolving width of Farallon slab. *Science* **329**, 316–319 (2010).
- Wortel, M. J. R. & Spakman, W. Subduction and slab detachment in the Mediterranean–Carpathian region. *Science* **290**, 1910–1917 (2000).
- McQuarrie, N. & Wernicke, B. An animated tectonic reconstruction of southwestern North America since 36 Ma. *Geosphere* **1**, 147–172 (2005).
- Priest, G. R. Volcanic and tectonic evolution of the cascade volcanic arc, central Oregon. *J. Geophys. Res.* **95**, 19583–19599 (1990).
- Hirth, G. & Kohlstedt, D. Water in the oceanic upper mantle: implications for rheology, melt extraction and the evolution of the lithosphere. *Earth Planet. Sci. Lett.* **144**, 93–108 (1996).
- Camp, V. & Hanan, B. A plume-triggered delamination origin for the Columbia River basalt group. *Geosphere* **4**, 480–495 (2008).
- Carlson, R. W. Isotopic constraints on Columbia River flood basalt genesis and the nature of the subcontinental mantle. *Geochim. Cosmochim. Acta* **48**, 2357–2372 (1984).
- Dodson, A., Kennedy, B. M. & DePaolo, D. J. Helium and neon isotopes in the Imnaha Basalt, Columbia River basalt group: evidence for a Yellowstone plume source. *Earth Planet. Sci. Lett.* **150**, 443–451 (1997).
- Takahashi, E., Nakajima, K. & Wright, T. L. Origin of the Columbia River basalts: melting model of a heterogeneous mantle plume head. *Earth Planet. Sci. Lett.* **162**, 63–80 (1998).
- Duncan, R. A. A captured island chain in the coast range of Oregon and Washington. *J. Geophys. Res.* **87**, 10827–10837 (1982).
- Beck, M. E. Has the Washington–Oregon coast range moved northward? *Geology* **12**, 737–740 (1984).
- Geist, D. & Richards, M. Origin of the Columbia Plateau and Snake River plain: deflection of the Yellowstone plume. *Geology* **21**, 789–792 (1993).
- Humphreys, E. D., Dueker, K. G., Schutt, D. L. & Smith, R. B. Beneath Yellowstone: evaluating plume and nonplume models using teleseismic images of the upper mantle. *GSA Today* **10**, 1–6 (2000).
- Glen, J. & Ponce, D. Large-scale fractures related to inception of the Yellowstone hotspot. *Geology* **30**, 647–650 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. Carlson and D. Blackman for discussions. Computational resources were provided by XSEDE project EAR100021. L.L. was funded by the John Miles Fellowship and the Cecil and Ida Green Foundation. D.R.S. was supported in part by the G. Unger Vetlesen Foundation.

Author Contributions L.L. designed and performed the numerical models. Both authors contributed equally to idea development and result interpretation.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to L.L. (lil019@ucsd.edu).

METHODS

Model set-up. We used the three-dimensional spherical finite element code for mantle convection, CitcomS²⁸, to simulate Farallon subduction. The code solves for an incompressible Newtonian fluid within a regional spherical mantle shell. We adopted a mesh with $257 \times 257 \times 65$ nodes in latitude \times longitude \times depth, covering a physical domain of $60^\circ \times 100^\circ \times 2,760$ km, respectively. The mesh had variable grid spacing in all three dimensions such that resolution increased towards the centre and the surface, with the finest mesh grid having a block size of $12 \text{ km} \times 20 \text{ km} \times 7 \text{ km}$ beneath the western USA. We ensured that the model box was wide enough to avoid sidewall-induced artificial flows. In our case, the box had the nearest vertical boundary over 2,000 km away from any part of the western USA. We also tested the effect of box depth on mantle flow and found that as long as the depth is larger than 1,000 km, the resulting upper mantle structures become similar. So we chose a mantle depth of approximately 2,700 km.

Initial and boundary conditions. These are essential components of a data assimilation model. The initial condition is both observationally constrained and numerically tested. The starting time of 40 Myr is consistent with both Cascadia volcanism history²⁹ and present-day seismic tomography³⁰. Tests suggest that subduction starting from this initial time are more than enough to properly capture the Farallon evolution since the Miocene, which is the time of interest in this study. The boundary conditions including both the thermal structure and surface kinematics are based on a recent plate reconstruction³¹. We define the top thermal boundary layer in oceanic plates using time-dependent palaeo-seafloor ages, with the profile largely following a half-space cooling model. The plate motions are imposed as top velocity boundary conditions at every time step during the calculation.

Another important feature adopted in our calculation is a pseudo-free surface implemented as a 'sticky air' layer on top of a viscous fluid, which has been shown to be successful in obtaining realistic slab geometry in the vicinity of a subduction zone as found in laboratory experiments³². We converted this sticky-air function into a phase transformation within the uppermost two elements of the model such that oceanic plate within this layer gains some extra negative buoyancy, to mimic the lateral pressure gradient at a convergent boundary due to the low topography of the trench. This effectively mimics the zero-density 'sticky air' implementation because they both increase the lateral gradient of buoyancy that promotes the plate's tendency to sink asymmetrically at the subduction front.

Rheologic structures. We use both depth- and temperature-dependent viscosity. A four-layer viscosity structure is assumed, including lithosphere, asthenosphere, transition zone, and lower mantle. The viscosity magnitudes of these layers are subsequently constrained by predicting the present-day seismic tomography image¹⁰. Density changes due to the phase transformation across the 410-km and 660-km interfaces were also considered. Lateral variation of viscosity is achieved first from temperature dependence, assuming a Newtonian fluid. We did not use a nonlinear rheology because of existing controversies over the viscous strength of mantle and slabs. We circumvent this problem by searching for the effective viscosity of ambient mantle and slabs so that we can best predict the observed slab morphology at the present day across the entire upper mantle. Furthermore, the model also incorporated many sharp rheological features such as narrow plate boundaries (vertical at mid-ocean ridges and transform faults, and one-side dipping above down-going slabs), slab hinges (weak bending parts of the slab with a stronger core), a mantle wedge (a weak zone between the surface and the slab), and large rheological variations from the Basin and Range province to

the cratonic North America. In effect, the model can achieve a viscosity contrast of up to four orders of magnitude within a 100-km distance. We find that the resulting strong three-dimensional viscosity variation is essential for the generation of asymmetric subduction and a physically reasonable subducting slab whose hinge closely follows the position of trenches as a function of time.

Mid-Miocene slab tear. With the model set-up as described above, we find that a segmented present-day mantle structure can easily be reached¹⁰. However, to predict the exact structures as observed by tomography³³, we have to find the appropriate viscosities for both the ambient mantle and slabs. Interestingly, we find that, besides the best-fitting model, those runs whose viscosities are several times larger or smaller than the best-fitting ones still have the Farallon slab break at a similar time and location, even though their predicted present-day structures are entirely off. This suggests that the slab tear was not sensitive to the large-scale viscosity structure. The detailed viscosity profile of the subduction zone and mantle wedge, on the other hand, plays a major part in the behaviour of the subducting slab, including its smoothness, curvature and trench retreat rate. For the slab hinge to follow trench position through time, a critical constraint from observations, the subduction zone has to be weak. However, we did not find evidence that this fine-scale rheology structure controls the slab tear formation.

Other parameters that may control the slab tear include surface plate motions and sub-slab dynamic pressure. We found that the imposed surface kinematics cannot generate the slab tear without enough sub-slab dynamic pressure (removing the sticky-air function decreases the dynamic pressure under the slab): when plate motions slowed down and trench rollback increased during the Miocene as observed, the slab would simply roll back faster but still follow the trench position without breaking internally around the mid-Miocene (Supplementary Fig. 3). The sub-slab dynamic pressure, however, is more important. Because the dynamic pressure is an intrinsic property of the three-dimensional slab geometry, it cannot be entirely separated out for a sensitivity test in a multi-parameter model like the one we have. But we can test its effect by turning on and off the sticky air function whose existence increases the sub-slab dynamic pressure by acknowledging the effect of trench topography. We found that without the sticky air, the slab would never break around the mid-Miocene even with the slow-down of surface plates, and the predicted present-day mantle structures would never match the tomography image. With the sticky air, the excess sub-slab dynamic pressure associated with the shrinking width of the Farallon plate builds up with time, and eventually breaks the slab midway along the trench where the pressure is the greatest (Fig. 3, Supplementary Fig. 3).

28. Zhong, S., Zuber, M. T., Moresi, L. N. & Gurnis, M. The role of temperature dependent viscosity and surface plates in spherical shell models of mantle convection. *J. Geophys. Res.* **105**, 11063–11082 (2000).
29. Humphreys, E. Post-Laramide removal of the Farallon slab, western United States. *Geology* **23**, 987–990 (1995).
30. Sigloch, K., McQuarrie, N. & Nolet, G. Two-stage subduction history under North America inferred from multiple-frequency tomography. *Nature Geosci.* **1**, 458–462 (2008).
31. Müller, R. D., Sdrolias, M., Gaina, C. & Roest, W. R. Age, spreading rates and spreading asymmetry of the world's ocean crust. *Geochim. Geophys. Geosyst.* **9**, Q04006 (2008a).
32. Schmeling, H. *et al.* A benchmark comparison of spontaneous subduction models—towards a free surface. *Phys. Earth Planet. Inter.* **171**, 198–223 (2008).
33. Schmandt, B. & Humphreys, E. Complex subduction and small-scale convection revealed by body-wave tomography of the western United States upper mantle. *Earth Planet. Sci. Lett.* **297**, 435–445 (2010).

DNase I sensitivity QTLs are a major determinant of human expression variation

Jacob F. Degner^{1,2*}, Athma A. Pai^{1*}, Roger Pique-Regi^{1*}, Jean-Baptiste Veyrieras^{1,3}, Daniel J. Gaffney^{1,4}, Joseph K. Pickrell¹, Sherry De Leon⁴, Katelyn Michelini⁴, Noah Lewellen⁴, Gregory E. Crawford^{5,6}, Matthew Stephens^{1,7}, Yoav Gilad¹ & Jonathan K. Pritchard^{1,4}

The mapping of expression quantitative trait loci (eQTLs) has emerged as an important tool for linking genetic variation to changes in gene regulation^{1–5}. However, it remains difficult to identify the causal variants underlying eQTLs, and little is known about the regulatory mechanisms by which they act. Here we show that genetic variants that modify chromatin accessibility and transcription factor binding are a major mechanism through which genetic variation leads to gene expression differences among humans. We used DNase I sequencing to measure chromatin accessibility in 70 Yoruba lymphoblastoid cell lines, for which genome-wide genotypes and estimates of gene expression levels are also available^{6–8}. We obtained a total of 2.7 billion uniquely mapped DNase I-sequencing (DNase-seq) reads, which allowed us to produce genome-wide maps of chromatin accessibility for each individual. We identified 8,902 locations at which the DNase-seq read depth correlated significantly with genotype at a nearby single nucleotide polymorphism or insertion/deletion (false discovery rate = 10%). We call such variants ‘DNase I sensitivity quantitative trait loci’ (dsQTLs). We found that dsQTLs are strongly enriched within inferred transcription factor binding sites and are frequently associated with allele-specific changes in transcription factor binding. A substantial fraction (16%) of dsQTLs are also associated with variation in the expression levels of nearby genes (that is, these loci are also classified as eQTLs). Conversely, we estimate that as many as 55% of eQTL single nucleotide polymorphisms are also dsQTLs. Our observations indicate that dsQTLs are highly abundant in the human genome and are likely to be important contributors to phenotypic variation.

It is now well established that eQTLs are abundant in a wide range of cell types and in diverse organisms, and recent studies have implicated human eQTLs as being important contributors to phenotypic variation^{1–5}. However, the underlying regulatory mechanisms by which eQTLs affect gene expression remain poorly understood. One mechanism that may be important is when the alternative alleles at a particular single nucleotide polymorphism (SNP) lead to different levels of transcription factor binding or nucleosome occupancy at regulatory sites; this in turn may lead to allele-specific differences in transcription rates^{9–12}. In this study we used DNase-seq in a panel of 70 individuals and found that a large fraction of eQTLs are indeed probably caused by this type of mechanism.

DNase-seq is a genome-wide extension of the classical DNase I footprinting method^{13–15}. This assay identifies regions of chromatin that are accessible (or ‘sensitive’) to cleavage by the DNase I enzyme. Such regions are referred to as DNase I-hypersensitive sites (DHSs). DNase I sensitivity provides a precise, quantitative marker of regions of open chromatin and is well correlated with a variety of other markers of active regulatory regions including promoter-associated

and enhancer-associated histone marks. Furthermore, bound transcription factors protect the DNA sequence within a binding site from DNase I cleavage, often producing recognizable ‘footprints’ of decreased DNase I sensitivity^{13,15–17}.

We collected DNase-seq data for 70 HapMap Yoruba lymphoblastoid cell lines for which gene expression data and genome-wide genotypes were already available^{6–8}. We obtained an average of 39 million uniquely mapped DNase-seq reads per sample, providing individual maps of chromatin accessibility for each cell line (see Supplementary Information for all analysis details). Our data allowed us to characterize the distribution of DNase I cuts within individual hypersensitive sites at extremely high resolution. As expected, the DHSs coincided to a great extent with previously annotated regulatory regions, and DNase I sensitivity was positively correlated with the expression levels of nearby genes (Supplementary Figs 6 and 7). Overall, the locations of hypersensitive sites were highly correlated across individuals (Supplementary Information)¹¹.

We tested for genetic variants that affect local chromatin accessibility. To do this, we divided the genome into non-overlapping 100-base-pair (bp) windows, and then focused our analysis on the 5% of windows with the highest DNase I sensitivity (see Supplementary Information). For each individual we treated the number of DNase-seq reads in a given window, divided by the total number of mapped reads, as a quantitative trait that estimated the level of chromatin accessibility. We then tested for association between individual-specific DNase I sensitivity in each window and genotypes of all SNPs and insertions/deletions (indels) in a *cis*-candidate region of 40 kilobases (kb) centred on the target window.

Using this procedure, we identified associations between genotypes and inter-individual variation in DNase-seq read depth in 9,595 windows at a false discovery rate (FDR) of 10% (corresponding to 8,902 distinct DHSs, once we combined adjacent windows whose hypersensitivity data were associated with the same SNP or indel; Fig. 1a). We refer to these 8,902 loci as ‘DNase I sensitivity QTLs’, or dsQTLs, and show an example in Fig. 1c–f. We additionally considered a much smaller *cis*-candidate region of only 2 kb around each target window and found that most of the dsQTLs were detected within this smaller region (7,088 associated windows in 6,070 DHSs), suggesting that most dsQTLs lie close to the target DHS. In contrast, we found only weak evidence of *trans*-acting dsQTLs, probably because our experiment was underpowered for detecting these (Supplementary Information). For dsQTLs with enough DNase-seq reads overlapping the most significant SNP ($n = 892$), we confirmed that the fraction of reads carrying each allele in heterozygotes was well correlated with the dsQTL effect sizes (correlation coefficient $r = 0.72$, $P \ll 10^{-16}$; Fig. 1b).

We observed that dsQTLs typically affected chromatin accessibility for about 200–300 bp (Fig. 2a). Of the DHSs affected by dsQTLs, 77%

¹Department of Human Genetics, University of Chicago, Chicago, Illinois 60637, USA. ²Committee on Genetics, Genomics and Systems Biology, University of Chicago, Chicago, Illinois 60637, USA.

³BioMiningLabs, 69001 Lyon, France. ⁴Howard Hughes Medical Institute, University of Chicago, Chicago, Illinois 60637, USA. ⁵Institute for Genome Sciences and Policy, Duke University, Durham, North Carolina 27708, USA. ⁶Department of Pediatrics, Division of Medical Genetics, Duke University School of Medicine, Durham, North Carolina 27708, USA. ⁷Department of Statistics, University of Chicago, Chicago, Illinois 60637, USA.

*These authors contributed equally to this work.

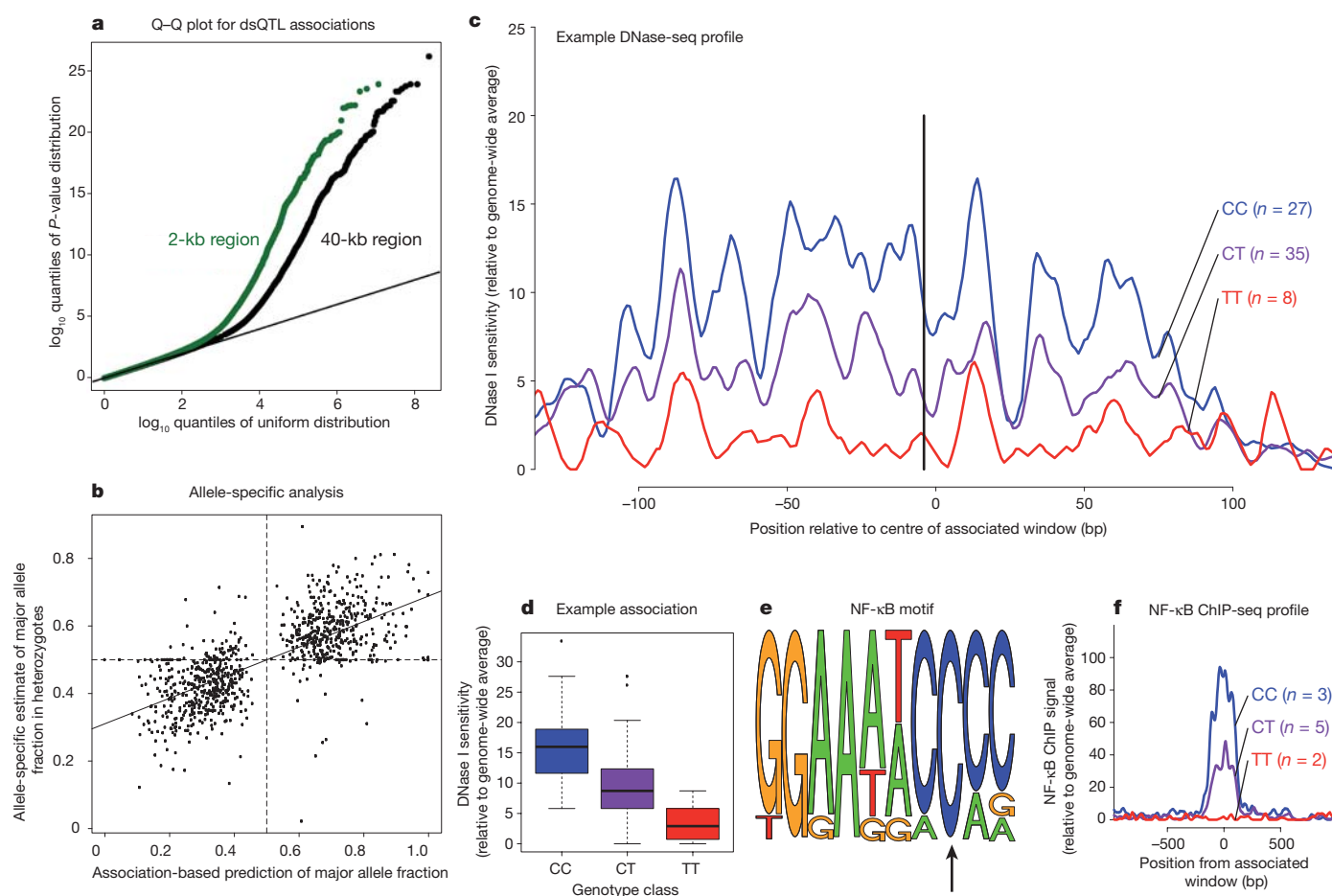


Figure 1 | Genome-wide identification of dsQTLs and a typical example.

a, Q-Q plots for all tests of association between DNase I cut rates in 100-bp windows, and variants within 2-kb (green) and 40-kb (black) regions centred on the target DHS windows. **b**, Allele-specific analysis of dsQTLs in heterozygotes. Plotted are the predicted (*x* axis) and observed (*y* axis) fractions of reads carrying the major allele based on the genotype means. **c**, Example of a

dsQTL (rs4953223). The black line indicates the position of the associated SNP. **d**, Box plot showing that rs4953223 is strongly associated with local chromatin accessibility ($P = 3 \times 10^{-13}$). **e**, The T allele, which is associated with low DNase I sensitivity, disrupts the binding motif of a previously identified NF- κ B-binding site at this location¹⁴. **f**, NF- κ B ChIP-seq data from ten individuals⁷ indicates a strong effect of this SNP on NF- κ B binding.

lie in chromatin regions previously predicted¹⁸ to be functional in lymphoblastoid cell lines: 41% in predicted enhancers, 26% in promoters, and 10% in insulators, even though those chromatin states together cover only 6.7% of the genome overall (and 38% of our hypersensitive sites).

We next studied the properties of *cis*-acting variants that generated dsQTLs, with the use of a Bayesian hierarchical model that accounted for the uncertainty about which sites are causal¹⁹ (Supplementary Information). This model obtained unbiased estimates of the average properties of causal sites even though, because of linkage disequilibrium, it was typically uncertain which site was causal for any individual dsQTL (Supplementary Information). As shown in Fig. 2b,c, most dsQTLs were generated by variants close to the target window. We estimate that 56% of the dsQTLs were due to variants that lay within the same DHSs and that 67% lay within 1 kb of the target window. dsQTLs that lay more than 1 kb from the target window were themselves significantly enriched in non-adjacent DHS windows (2.4-fold compared with matched random SNPs) and were often associated with changes in sensitivity in multiple non-adjacent DHS windows (Supplementary Fig. 15).

One intuitive mechanism for dsQTLs is that these may be caused by variants that strengthen or weaken individual transcription factor binding sites, thereby changing transcription factor affinity and local nucleosome occupancy^{20–22} and hence DNase I cut rates. Consistent with this model, an aggregated plot of DNase I sensitivity at dsQTLs showed a distinct drop in chromatin accessibility around putatively

causal SNPs that was reminiscent of transcription factor binding footprints, especially in the genotypes associated with high sensitivity^{15–17}.

To test the importance of disruption of transcription factor binding sites as a mechanism underlying dsQTLs, we again turned to the Bayesian hierarchical model. We used the union of all published footprint locations in lymphoblastoid cell lines^{16,17} and a set of footprints that we identified from the DNase-seq data reported in this study (Supplementary Methods). Analysis using the hierarchical model indicated a 3.6-fold enrichment of dsQTLs within transcription factor binding footprints ($P \ll 10^{-16}$), controlling for the overall enrichment within DHSs. In addition, the allele associated with a higher score of the position weight matrix is typically associated with higher chromatin accessibility ($P \ll 10^{-16}$), which is consistent with the expectation that higher transcription factor binding affinity leads to more open chromatin (Fig. 2d). Of the dsQTLs that fell within DNase-seq footprints tied to specific transcription factor motifs (using CENTIPEDE¹⁷), CCCTC binding factor (CTCF), cAMP-response element (CRE) and interferon-stimulated response element (ISRE) were the most enriched, whereas MADS box transcription enhancer factor 2 (MEF2) was significantly depleted.

To further understand the functional consequences of dsQTLs, we examined ChIP-seq data for nine transcription factors collected by the ENCODE Project in one or more lymphoblastoid cell lines^{10,23}. Overall, the alleles that were associated with increased DNase I sensitivity were highly associated with increased transcription factor

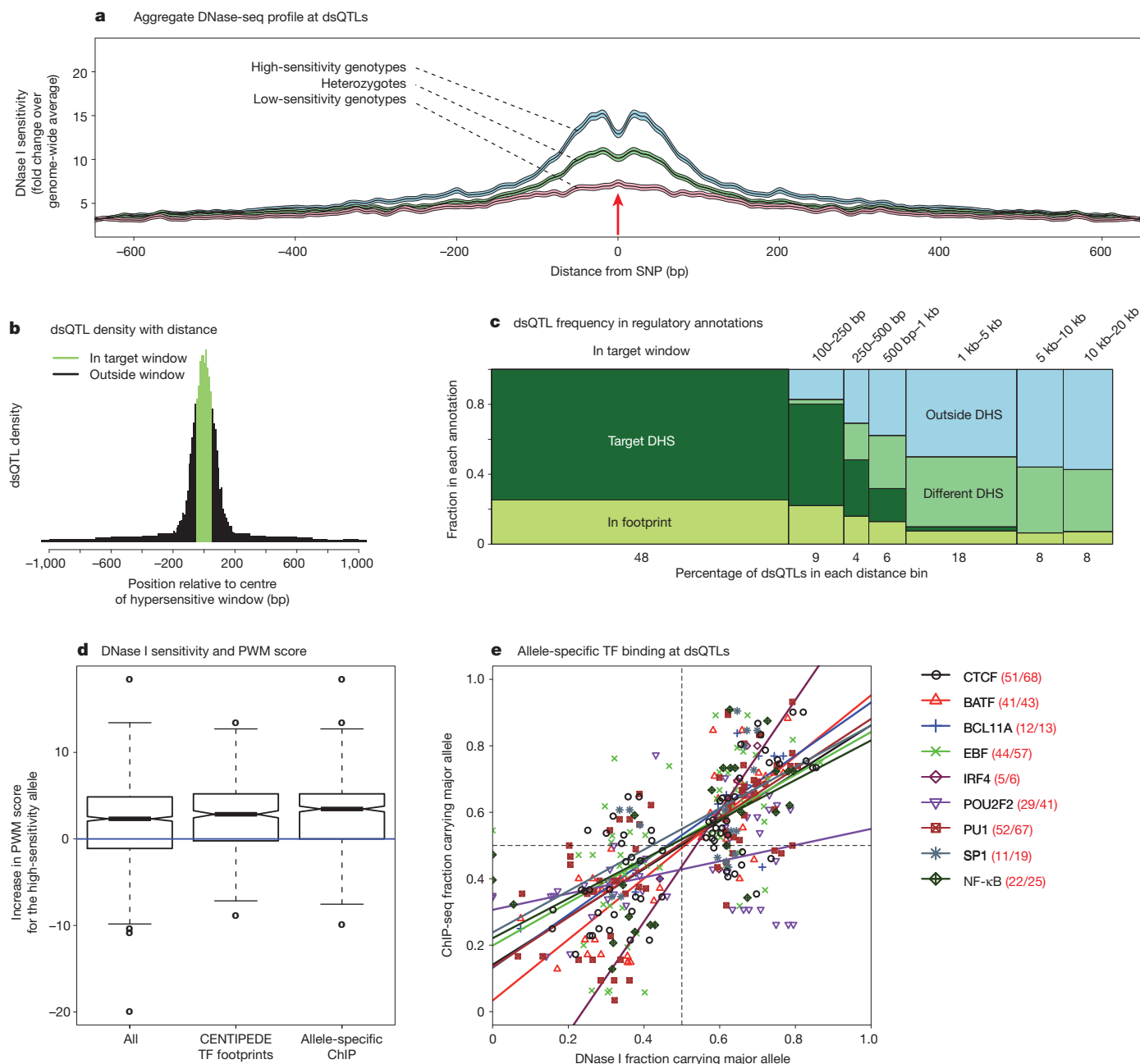


Figure 2 | Properties of dsQTLs. **a**, Aggregated plot of DNase I sensitivity for high-confidence dsQTLs that lie within the target DHS. Individuals were separated into the high-sensitivity (blue), heterozygote (green), and low-sensitivity (red) classes. The shading indicates the bootstrap 95% confidence intervals. **b**, The peak density of dsQTLs is very tightly focused around the target DHS window. **c**, Total fraction of *cis*-dsQTLs that fall into different categories of distance from the target window (*x* axis) and different annotations (*y* axis). The total area of each rectangle is proportional to the estimated number of dsQTLs in that category. **d**, Box plot showing distribution of position weight matrix (PWM) score differences between high-sensitivity and low-sensitivity dsQTL alleles, respectively. Notches indicate 95% confidence intervals for

binding ($P < 10^{-16}$; Fig. 2e), indicating that dsQTLs are strong predictors of changes in occupancy by a range of DNA-binding proteins.

Given that dsQTLs produce sequence-specific changes in chromatin accessibility and, frequently, changes in transcription factor binding, we speculated that a fraction of the dsQTL variants might also affect expression levels of nearby genes. We examined this by testing for associations between the most significant variant at each of the dsQTLs detected by using the 2 kb window size and expression

median. **e**, The *x* axis shows the fraction of sequence reads predicted to carry the major allele based on the DNase I genotype means; the *y* axis shows the observed fraction in ChIP-seq data. The lines show the regression fits for each factor separately; the numbers in the key show the fraction of sites that are in a concordant direction for each factor. CTCF, CTCF binding factor; BATF, basic leucine zipper transcription factor; BCL11A, B-cell CLL/lymphoma 11A zinc-finger protein; EBF, early B-cell factor 1; IRF4, interferon regulatory factor 4; POU2F2, POU class 2 homeobox 2; PU1, proviral integration oncogene *spi1*; SP1, Sp1 transcription factor; NF-κB, nuclear factor of κ light polypeptide gene enhancer in B-cells 1.

levels of nearby genes (that is, genes with transcription start sites (TSSs) within 100 kb) estimated by sequencing RNA from the same cell lines⁸. Using this approach, we found that 16% of dsQTL SNPs were also significantly associated with variation in expression levels of at least one nearby gene (FDR = 10%). This represents a huge enrichment over random expectation (450-fold, $P \ll 10^{-16}$; Fig. 3). One example of a joint dsQTL-eQTL is illustrated in Fig. 3a, in which a SNP disrupts an ISRE located in the first intron of the *SLFN5* gene,

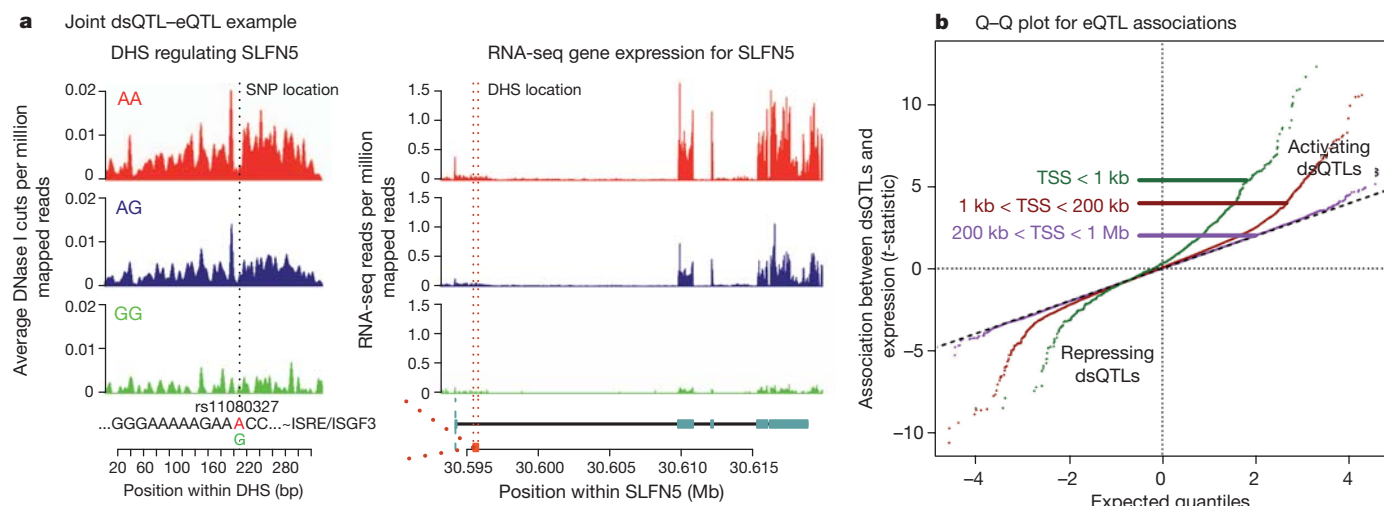


Figure 3 | Relationship between dsQTLs and eQTLs. **a**, Example of a dsQTL SNP that is also an eQTL for the gene *SLFN5*. The SNP disrupts an interferon-sensitive response element, thereby changing local chromatin accessibility within the first intron of *SLFN5*. Expression of *SLFN5* has been shown to be inducible by interferon α in melanoma cell lines. DNase-seq (left) and RNA-seq

(right) measurements from DNase-seq and RNA-seq are plotted, stratified by genotype at the putative causal SNP. **b**, Q-Q plot of the *t*-statistic for association with gene expression changes (eQTL) of dsQTL SNPs. The sign of the eQTL *t*-statistic is with respect to the genotype that increases DNase sensitivity.

leading to both a strong dsQTL and an eQTL for *SLFN5*. Conversely, out of 1,271 eQTLs detected by using RNA-seq data from these cell lines⁸, 23% of the most significant SNPs were also dsQTLs (FDR = 10%). Using the method in ref. 24 for estimating the proportion of tests in which the null hypothesis is false (while accounting for incomplete power), we estimate that 55% of the most significant eQTL SNPs are also dsQTLs and that 39% of the dsQTLs are also eQTLs. dsQTLs are therefore a major mechanism by which genetic variation may affect gene expression levels.

We observed that for most (70%) of the joint dsQTL-eQTLs, the allele that was associated with increased chromatin accessibility was also associated with increased gene expression levels (Fig. 3b). Because higher DNase I sensitivity generally correlates with higher transcription factor occupancy, this suggests that transcription factors that are bound to DHSs usually act as enhancers. CRE-box and ETS-box were the most enriched motifs among repressors and enhancers, respectively. The dsQTLs that were also eQTLs (FDR = 10%) were highly enriched around the TSSs of the target genes: for 23% of the joint dsQTL-eQTLs, the associated DHS was within 1 kb of the TSS, and for 39% it was within 10 kb (Fig. 4a). This is consistent with previous work showing strong clustering of eQTLs around TSSs^{19,25,26}. Nonetheless, there was a significant signal of long-range regulation as far as 100 kb. In addition, 14% of the joint dsQTL-eQTLs were significant eQTLs for two or more genes, suggesting that some regulatory regions affect more than one gene.

We sought to identify additional factors that might influence whether a dsQTL regulates gene expression of nearby genes, while controlling for the very strong effect of distance from TSS (Fig. 4b). We observed that a dsQTL was more likely to be an eQTL for the gene with the nearest TSS (1.6-fold, $P = 3 \times 10^{-4}$) and was more likely to be an eQTL if it was located within the transcribed region of the gene (2.7-fold, $P = 2 \times 10^{-9}$). Further, a dsQTL was 2.6-fold more likely to be an eQTL if it was associated with a DHS that overlapped a DNA methylation QTL²⁷ ($P = 4 \times 10^{-4}$), and showed a 2.4-fold increase if the associated DHS overlapped a RNA polymerase II ChIP-seq peak¹⁰ ($P = 4 \times 10^{-4}$). Conversely, a dsQTL was significantly less likely to be an eQTL for a gene if an active binding site for the insulator protein CTCF¹⁷ lay between the dsQTL and the gene's TSS (2.4-fold decrease, $P = 10^{-12}$). Finally, the presence of the enhancer mark P300 (from ENCODE ChIP-seq data²⁸) in the dsQTL window increased the probability that a distal dsQTL (TSS > 1.5 kb) was an eQTL (1.7-fold, $P = 10^{-5}$).

We have shown here that common genetic variants affect chromatin accessibility at thousands of hypersensitive regions across the human genome. The putative causal variants most often lie within or very near the hypersensitive regions, and frequently act by changing the binding affinity of transcription factors. Mapping of dsQTLs provides

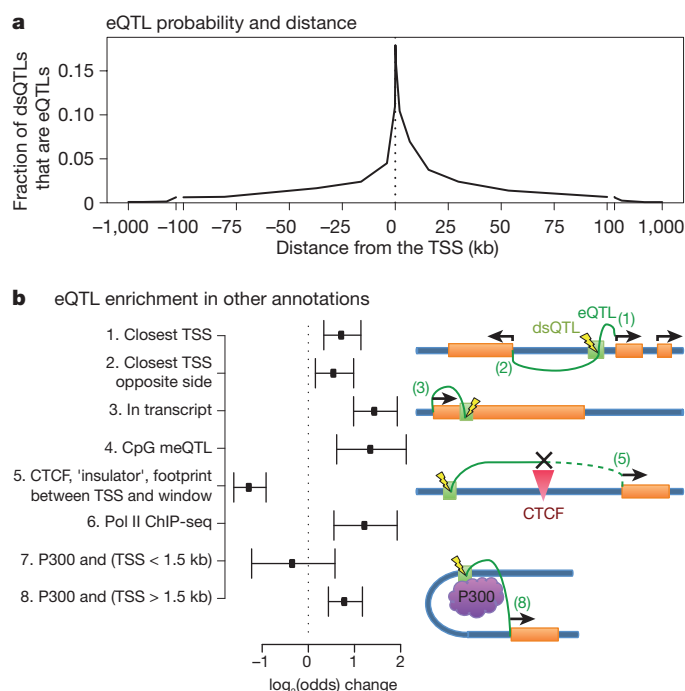


Figure 4 | Relationship between dsQTLs and eQTLs. **a**, Most joint dsQTL-eQTLs lie close to the gene TSS. **b**, Effect of various factors on the log odds that a given dsQTL is also an eQTL, while controlling for the strong distance relationship observed in **a**. In annotations (1) and (2) we do not consider the direction of transcription. In annotations (6–8) ChIP-seq is measured on the dsQTL window. In annotations (4) and (6), 'meQTL' refers to a dsQTL that is also associated with methylation levels of a nearby CpG site²⁷ and 'Pol II' refers to the presence of an RNA polymerase II ChIP-seq peak overlapping the DHS associated with the dsQTL²³. One of the most significant annotations in delineating the regulatory regions is defined by the presence of the CTCF insulator element, which decreases 2.4-fold the probability that a dsQTL is an eQTL. Error bars represent 95% confidence intervals.

a powerful tool for detecting potentially functional changes in a variety of different types of regulatory element, and roughly 50% of eQTLs are also dsQTLs. Furthermore, analysis of significantly associated SNPs from genome-wide association studies additionally implicates some of these dsQTLs as potentially underlying a variety of genome-wide association study hits (Supplementary Information). Changes in chromatin accessibility may be a major mechanism linking genetic variation to changes in gene regulation and, ultimately, organismal phenotypes.

METHODS SUMMARY

DNase-seq libraries were created as described previously²⁹, with small modifications. Each library was sequenced on at least two lanes of an Illumina GAIIx. Resulting 20-bp sequencing reads were mapped to the human genome sequence (hg18) using an algorithm that we designed specifically to eliminate mappability biases between sequence variants. We divided the genome into 100-bp windows and selected the top 5% in terms of total DNase I sensitivity. DNase I sensitivity for each individual in each window was normalized by the total number of mapped reads for that individual. For QTL mapping, the data were further rescaled within and across individuals, and we adjusted the data for an observed individual \times GC interaction, as well as for the top four principal components of the DNase I sensitivity matrix. Genotypes for all available SNPs and indels were obtained from HapMap and 1,000 Genomes data and imputed where necessary^{6,7,30}. We performed DNase-seq association mapping by regressing the adjusted sensitivity in each window against the genotypes at variants in a 40-kb region centred on each DHS. As validation, we used our DNase-seq reads as well as ChIP-seq reads and DNase-seq reads from ENCODE to confirm that allele-specific reads spanning heterozygous sites at dsQTLs were consistent with the association analysis. We also used RNA-seq data from the same cell lines⁸ to study the links between dsQTLs and eQTLs. Finally, we explored the properties of dsQTLs that made them more or less likely to influence gene expression by fitting a logistic model on all dsQTLs, where the eQTL status of each dsQTL–eQTL test was modelled as a function of distance from the TSS and a variety of other annotations. For full details of all methods see Supplementary Information.

Received 23 June; accepted 15 December 2011.

Published online 5 February 2012.

- Brem, R. B., Yvert, G., Clinton, R. & Kruglyak, L. Genetic dissection of transcriptional regulation in budding yeast. *Science* **296**, 752–755 (2002).
- Cheung, V. G. *et al.* Mapping determinants of human gene expression by regional and genome-wide association. *Nature* **437**, 1365–1369 (2005).
- Nicolae, D. L. *et al.* Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet.* **6**, e1000888 (2010).
- Nica, A. C. *et al.* Candidate causal regulatory effects by integration of expression QTLs with complex trait genetic associations. *PLoS Genet.* **6**, e1000895 (2010).
- Lango Allen, H. *et al.* Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* **467**, 832–838 (2010).
- Frazer, K. A. *et al.* A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851–861 (2007).
- Durbin, R. M. *et al.* A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).
- Pickrell, J. K. *et al.* Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature* **464**, 768–772 (2010).
- Gaulton, K. J. *et al.* A map of open chromatin in human pancreatic islets. *Nature Genet.* **42**, 255–259 (2010).
- Kasowski, M. *et al.* Variation in transcription factor binding among humans. *Science* **328**, 232–235 (2010).
- McDaniell, R. *et al.* Heritable individual-specific and allele-specific chromatin signatures in humans. *Science* **328**, 235–239 (2010).
- Zheng, W., Zhao, H., Mancera, E., Steinmetz, L. M. & Snyder, M. Genetic analysis of variation in transcription factor binding in yeast. *Nature* **464**, 1187–1191 (2010).
- Galas, D. & Schmitz, A. DNase footprinting: a simple method for the detection of protein-DNA binding specificity. *Nucleic Acids Res.* **5**, 3157–3170 (1978).
- Boyle, A. P. *et al.* High-resolution mapping and characterization of open chromatin across the genome. *Cell* **132**, 311–322 (2008).
- Hesselberth, J. R. *et al.* Global mapping of protein-DNA interactions *in vivo* by digital genomic footprinting. *Nature Methods* **6**, 283–289 (2009).
- Boyle, A. P. *et al.* High-resolution genome-wide *in vivo* footprinting of diverse transcription factors in human cells. *Genome Res.* **21**, 456–464 (2011).
- Pique-Regi, R. *et al.* Accurate inference of transcription factor binding from DNA sequence and chromatin accessibility data. *Genome Res.* **21**, 447–455 (2011).
- Ernst, J. *et al.* Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**, 43–49 (2011).
- Veyrieras, J. B. *et al.* High-resolution mapping of expression-QTLs yields insight into human gene regulation. *PLoS Genet.* **4**, e1000214 (2008).
- Mirny, L. A. Nucleosome-mediated cooperativity between transcription factors. *Proc. Natl Acad. Sci. USA* **107**, 22534–22539 (2010).
- Wasson, T. & Hartemink, A. J. An ensemble model of competitive multi-factor binding of the genome. *Genome Res.* **19**, 2101–2112 (2009).
- Raveh-Sadka, T., Levo, M. & Segal, E. Incorporating nucleosomes into thermodynamic models of transcription regulation. *Genome Res.* **19**, 1480–1496 (2009).
- Myers, R. M. *et al.* A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol.* **9**, e1001046 (2011).
- Storey, J. D., Taylor, J. E. & Siegmund, D. Strong control, conservative point estimation, and simultaneous conservative consistency of false discovery rates: a unified approach. *J. R. Stat. Soc., B* **66**, 187–205 (2004).
- Dixon, A. L. *et al.* A genome-wide association study of global gene expression. *Nature Genet.* **39**, 1202–1207 (2007).
- Stranger, B. E. *et al.* Population genomics of human gene expression. *Nature Genet.* **39**, 1217–1224 (2007).
- Bell, J. T. *et al.* DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. *Genome Biol.* **12**, R10 (2011).
- Visel, A. *et al.* ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457**, 854–858 (2009).
- Song, L. & Crawford, G. E. DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. *Cold Spring Harb. Protocols*. doi:10.1101/pdb.prot5384 (2010).
- Guan, Y. & Stephens, M. Practical issues in imputation-based association mapping. *PLoS Genet.* **4**, e1000279 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank members of the Pritchard, Przeworski, Stephens and Gilad laboratories for many helpful comments or discussions, and the ENCODE Project for publicly available ChIP-seq data. This work was supported by grants from the National Institutes of Health to Y.G. (HG006123) and J.K.P. (MH084703 and MH090951), by the Howard Hughes Medical Institute, by the Chicago Fellows Program (to R.P.R.), by the American Heart Association (to A.A.P.), and by the NIH Genetics and Regulation Training grant (A.A.P. and J.F.D.).

Author Contributions A.A.P. led the data collection with assistance from S.D.L., K.M. and N.L. The data analysis was performed jointly by J.F.D. and R.P.R., with contributions from A.A.P., J.B.V., D.J.G. and J.K.P. G.E.C. and M.S. provided technical assistance and discussion of methods and results. The manuscript was written by J.F.D., A.A.P., R.P.R., Y.G. and J.K.P. The project was jointly supervised by Y.G. and J.K.P.

Author Information All raw data and tables of all dsQTLs are deposited in GEO under accession number GSE31388 and at <http://eqtl.uchicago.edu>. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.K.P. (pritch@uchicago.edu) or Y.G. (gilad@uchicago.edu).

Extrathymically generated regulatory T cells control mucosal T_H2 inflammation

Steven Z. Josefowicz^{1,2*}, Rachel E. Niec^{1*}, Hye Young Kim³, Piper Treuting⁴, Takatoshi Chinen^{1,5}, Ye Zheng⁶, Dale T. Umetsu³ & Alexander Y. Rudensky¹

A balance between pro- and anti-inflammatory mechanisms at mucosal interfaces, which are sites of constitutive exposure to microbes and non-microbial foreign substances, allows for efficient protection against pathogens yet prevents adverse inflammatory responses associated with allergy, asthma and intestinal inflammation¹. Regulatory T (T_{reg}) cells prevent systemic and tissue-specific autoimmunity and inflammatory lesions at mucosal interfaces. These cells are generated in the thymus (tT_{reg} cells) and in the periphery (induced (i)T_{reg} cells), and their dual origin implies a division of labour between tT_{reg} and iT_{reg} cells in immune homeostasis. Here we show that a highly selective blockage in differentiation of iT_{reg} cells in mice did not lead to unprovoked multi-organ autoimmunity, exacerbation of induced tissue-specific autoimmune pathology, or increased pro-inflammatory responses of T helper 1 (T_H1) and T_H17 cells. However, mice deficient in iT_{reg} cells spontaneously developed pronounced T_H2-type pathologies at mucosal sites—in the gastrointestinal tract and lungs—with hallmarks of allergic inflammation and asthma. Furthermore, iT_{reg}-cell deficiency altered gut microbial communities. These results suggest that whereas T_{reg} cells generated in the thymus appear sufficient for control of systemic and tissue-specific autoimmunity, extrathymic differentiation of T_{reg} cells affects commensal microbiota composition and serves a distinct, essential function in restraint of allergic-type inflammation at mucosal interfaces.

Exquisitely balanced control mechanisms operating at mucosal sites are able to accommodate potent immune defences and the need to prevent tissue damage resulting from inflammatory responses caused by commensal microorganisms, food and environmental antigens, allergens, and noxious substances¹.

Prominent among multiple regulatory lymphoid and myeloid cell subsets operating at environmental interfaces are Foxp3⁺ T_{reg} cells. Genetic deficiency in Foxp3 (forkhead box P3, a key transcription factor specifying T_{reg} cell differentiation) leads to paucity of Foxp3⁺ T_{reg} cells and consequent generalized lympho- and myelo-proliferative syndrome, featuring sharply augmented serum IgE levels, production of T_H1, T_H2 and T_H17 cytokines, and widespread tissue inflammation². Foxp3 can be induced in thymocytes in response to T-cell receptor (TCR) and CD28 stimulation, and IL-2. In addition, Foxp3 can be upregulated upon TCR stimulation of mature peripheral CD4⁺ T cells in the presence of tumour growth factor β (TGF β) in a manner dependent on an intronic *Foxp3* enhancer CNS1 (refs 3–5). Inflammatory cytokines and potent co-stimulatory signals antagonize the peripheral induction of Foxp3, and retinoic acid augments Foxp3 induction through mitigating inflammatory cytokine production and through cell intrinsic mechanisms^{1,6–8}. Although differing in their sites of generation, tT_{reg} and iT_{reg} cells are comingled in the secondary lymphoid organs and non-lymphoid tissues once mature, and their relative contributions to the total population of T_{reg} cells and their

specific roles in control of various aspects of immune homeostasis and microbial colonization in normal animals has remained unexplored.

Our recent investigation⁵ showed that CNS1, which contains binding sites for transcription factors (NFAT, Smad3 and RAR/RXR) downstream of three signalling pathways implicated in iT_{reg} cell generation^{4,8} (Supplementary Fig. 1), is critical for TGF β -dependent induction of Foxp3, but has no apparent role in tT_{reg} differentiation or maintenance of Foxp3 expression. This observation suggested that CNS1 activity represents a dedicated genetic determinant for the differentiation of iT_{reg} cells, and its deficiency in mice provides a unique means to evaluate the function of these cells *in vivo*. Our initial characterization of CNS1[−] mice and littermates maintained on a 129/B6 genetic background failed to reveal disease phenotypes. Because mixed genetic backgrounds frequently mask adverse phenotypes or make them highly variable, to understand iT_{reg} function *in vivo* we backcrossed CNS1 mice onto the B6 background (Supplementary Fig. 2).

First, we sought to ascertain that on the B6 genetic background CNS1 is dispensable for tT_{reg} cell generation but critical for generation of iT_{reg} cells. Two recent studies established a role for TGF β signalling in tT_{reg} cell differentiation in neonates^{9,10}. Thus, to exclude the possibility that CNS1 deficiency adversely affects generation of Foxp3⁺ T cells in the neonatal thymus, we examined the Foxp3⁺ T_{reg} cell population in heterozygous female CNS1^{WT/−} mice. As Foxp3 is encoded on the X chromosome and is subject to random X-chromosome inactivation, characterization of female CNS1^{WT/−} mice allows for comparison of CNS1[−] and CNS1^{WT} T_{reg} cells in a competitive environment. In neonatal female CNS1^{WT/−} mice, CNS1[−] cells constituted, on average, one-half of the thymic Foxp3⁺ cell population (Fig. 1a). Additionally, neonatal CNS1[−] hemizygous and control males harboured comparable numbers of Foxp3⁺ thymocytes (Supplementary Fig. 3). Therefore, tT_{reg} differentiation is independent of CNS1. In contrast, CNS1[−] naive CD4 T cells showed severely impaired induction of Foxp3 *in vitro* (Fig. 1b). Analyses of heterozygous female CNS1^{WT/−} mice and transfer of CNS1[−] or CNS1^{WT} T_{reg} cells into lymphopenic recipients demonstrated that the ability of T_{reg} cells to accumulate and proliferate in various tissues was unperturbed in the absence of CNS1 (Supplementary Fig. 4). Furthermore, CNS1 deficiency did not affect suppressor activity of tT_{reg} cells (assessed using *in vitro* suppression assays and adoptive transfers of Foxp3-deficient effector T cells with predominantly tT_{reg}-containing Foxp3⁺ cells isolated from 4-week-old CNS1[−] and CNS1^{WT} mice into lymphopenic recipients (Supplementary Fig. 5)). Likewise, CNS1 ablation did not negatively affect maintenance of Foxp3 expression and overall function of NFAT, TGF β and retinoic acid signalling pathways in these cells (Supplementary Fig. 5 and data not shown). To assess how the deficiency in iT_{reg} cell generation affects the size of the peripheral T_{reg} cell compartment, we analysed T_{reg} cell frequencies in various tissues throughout the lifespan of mice. CNS1[−] mice failed to exhibit a progressive age-dependent increase in Foxp3⁺

¹Howard Hughes Medical Institute and Immunology Program, Sloan Kettering Institute, New York, New York 10021, USA. ²Laboratory of Chromatin Biology and Epigenetics, The Rockefeller University, New York, New York 10065, USA. ³Division of Immunology, Children's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA. ⁴Department of Comparative Medicine, and Histology and Imaging Core, School of Medicine, University of Washington, Seattle, Washington 98195, USA. ⁵Department of Microbiology and Immunology, Keio University School of Medicine, Tokyo 160-8582, Japan. ⁶Norris Foundation Laboratories for Immunobiology and Microbial Pathogenesis, Salk Institute for Biological Studies, La Jolla, California 92037, USA.

*These authors contributed equally to this work.

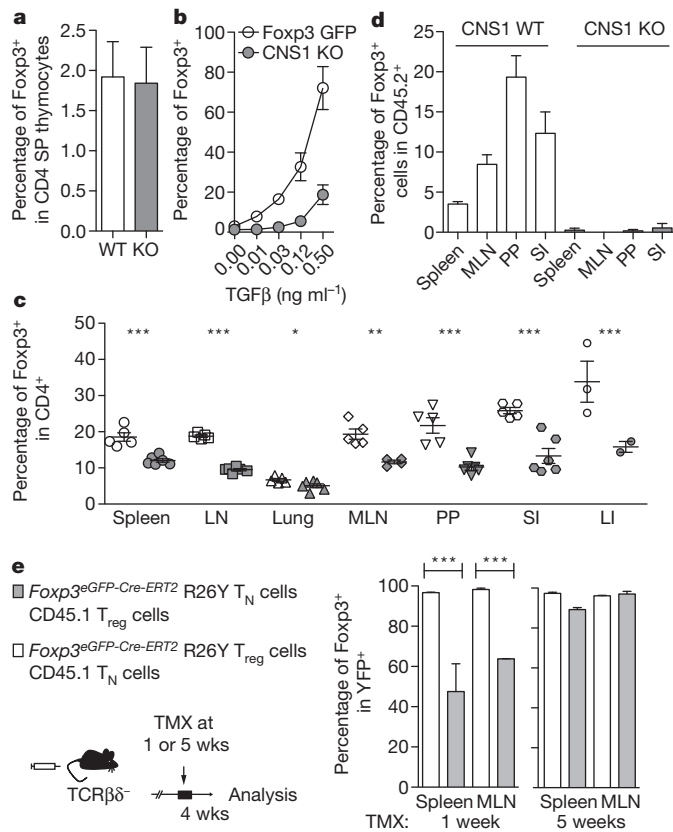


Figure 1 | Impaired iT_{reg} cell generation and altered composition of the peripheral T_{reg} cell population in CNS1-deficient mice. **a**, Relative contribution of CNS1⁻ (GFP⁺) and CNS1^{WT} (GFP⁻) cells to the Foxp3⁺ thymocyte subset in 4-day-old CNS1^{WT/-} female mice. SP, single positive. **b**, Induction of Foxp3 in Foxp3⁻ T_N (naive) cells FACS sorted from CNS1⁻ (knockout, KO) or Foxp3^{GFP} mice stimulated *in vitro* with TGFβ, IL-2 anti-CD3 and anti-CD28. **c**, Percentage of Foxp3⁺ cells (of CD4⁺) in the spleen, lymph node (LN), mesenteric lymph nodes (MLN), Peyer's patches (PP) and cells from the small and large intestine lamina propria (SI and LI) of 6–9 month old CNS1⁻ or control mice. **d**, Percentage of transferred (CD45.2⁺) CNS1⁻ or CNS1^{WT} CD25⁻ CD44^{low} CD45.2⁺ OTII⁺ cells that induced Foxp3 following administration of OVA in water for 6 days. **e**, Stability of Foxp3 expression in iT_{reg} cells. FACS sorted GFP⁺ or GFP⁻ cells from Foxp3^{GFP}-Cre-ERT2 mice were transferred with GFP⁺ or GFP⁻ cells, respectively, from CD45.1 Foxp3^{GFP} mice into TCRβδ-deficient recipients. Mice received tamoxifen (TMX) at 1 (left) or 5 weeks (right) after transfer and stability of Foxp3 expression among YFP-labelled cells was assessed after 4 weeks. All data are representative of two or more independent experiments with $n \geq 3$. Error bars, s.d.; * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, as calculated by Students' *t*-test.

cell frequencies observed in wild-type littermates (Fig. 1c and Supplementary Fig. 6). By 6–8 months of age, CNS1⁻ mice contained markedly fewer Foxp3⁺ cells in comparison to control animals, with most prominent differences in mesenteric lymph nodes, Peyer's patches, and small and large intestine lamina propria, sites known to support iT_{reg} cell generation¹¹. This trend was not the result of expression of a Foxp3–GFP fusion protein in CNS1⁻ mice, because age-matched CNS1^{WT} Foxp3–GFP and littermate control CNS1^{WT} mice expressing unmodified Foxp3 protein exhibited similar age-dependent increases in T_{reg} cell frequencies (Supplementary Fig. 6).

To assess the extent of impairment of peripheral generation of T_{reg} cells *in vivo*, we examined Foxp3 induction in antigen-specific naive T cells upon exposure to ingested 'non-self' antigen¹². Ovalbumin (OVA)-specific OT-II⁺ TCR-transgenic Foxp3⁻ (GFP⁻) T_{reg} cells from CNS1⁻ or Foxp3^{GFP} mice were transferred into CD45.1⁺ lymphoreplete recipients followed by *ad libitum* administration of OVA in drinking water. We failed to detect Foxp3 induction in CNS1-deficient cells, whereas up to 20% of transferred OT-II T cells

from control Foxp3^{GFP} mice induced Foxp3 upon exposure to cognate antigen in the intestinal tract (Fig. 1d and Supplementary Fig. 7). These results were in agreement with a marked impairment in Foxp3 induction in polyclonal CNS1-deficient Foxp3⁻ T cells *in vitro*, which was most severe at lower, more physiologically relevant concentrations of TGFβ (Fig. 1b). Together these data indicate that iT_{reg} cells have a stringent requirement for CNS1 for their differentiation.

Recent studies showed a limited TCR-dependent clonal niche for tT_{reg} cell differentiation and peripheral maintenance^{13–15}. The sustained numerical impairment in the peripheral T_{reg} cell populations in CNS1-deficient mice suggests that tT_{reg} cells fail to fill the 'void' in the peripheral T_{reg} cell pool, left by iT_{reg} cell deficiency. This observation combined with largely non-overlapping TCR repertoires of tT_{reg} and iT_{reg} cells suggests that iT_{reg} and tT_{reg} cells occupy distinct 'niches'¹⁶. To test this notion we co-transferred CNS1⁻ (tT_{reg} cells) or CNS1^{WT} T_{reg} cells (iT_{reg} + tT_{reg}) from aged mice with CNS1-sufficient naive CD45.1⁺ Foxp3⁻ CD4⁺ T cells into lymphopenic recipients. We observed more efficient Foxp3 induction in CD45.1⁺ CD4⁺ T cells upon co-transfer with CNS1⁻ T_{reg} cells (tT_{reg} cells), indicating that in lymphopenic recipients the *de novo* generation of iT_{reg} cells is markedly more efficient in the absence of pre-existing iT_{reg} cells (Supplementary Fig. 8). These data also imply the existence of a stable iT_{reg} cell subset in normal mice. However, the dynamics and stability of Foxp3 expression has been a controversial issue, with a number of studies favouring unstable Foxp3 expression in iT_{reg} cells^{17–19}. Thus, we next employed genetic fate mapping using inducible Cre recombinase expressed in a T_{reg}-specific manner (Foxp3^{GFP}-Cre-ERT2) and a Rosa26–YFP recombination reporter allele (R26Y)²⁰ to determine if iT_{reg} cells generated *in vivo* are able to acquire stable Foxp3 expression and, thus, have the capacity to contribute to the stable T_{reg} cell compartment.

Double-sorted naive CD45.2⁺ Foxp3⁻ YFP⁻ CD4 T cells from Foxp3^{GFP}-Cre-ERT2 R26Y mice were transferred together with congenically marked CD45.1⁻ Foxp3⁺ T_{reg} cells into lymphopenic recipient mice. Foxp3 expression within the population of tagged YFP⁺ cells generated from YFP⁻ Foxp3⁻ precursors was assessed four weeks after treatment of recipient mice with tamoxifen, which was administered early (one week) and late (five weeks) following cell transfer. Approximately half of the newly generated YFP-tagged iT_{reg} cells lost Foxp3 expression, whereas 'mature' iT_{reg} cells tagged at a later time point displayed remarkable stability (>90% Foxp3⁺ cells among YFP⁺ cells), comparable to that of transferred peripheral T_{reg} cells (Fig. 1e and Supplementary Fig. 9). Together these data indicate that iT_{reg} cells have a stringent requirement for CNS1 for their differentiation, accumulate throughout life, and occupy a sizable fraction of the stable peripheral T_{reg} cell compartment.

CNS1⁻ mice on the B6 genetic background displayed neither early-onset late-onset systemic autoimmunity nor spontaneous widespread tissue lesions nor severe morbidity associated with systemic T_{reg} cell deprivation (data not shown). However, it was possible that iT_{reg} cell deficiency may exacerbate initial or late stages of provoked tissue-specific autoimmune pathology directed against a self-antigen. To address this question, we induced experimental autoimmune encephalomyelitis (EAE) in CNS1-deficient or littermate control mice through immunization with myelin oligodendrocyte glycoprotein (MOG) peptide. The onset, severity and remission of disease were indistinguishable, and no detectable differences were observed in T_{reg} cell subsets in the brain in these two groups of mice (Supplementary Fig. 10). Although it will be important to evaluate the role of iT_{reg} cells in additional models of induced autoimmunity, these results indicate that tT_{reg} cells are largely sufficient for control of tolerance to self-antigens and that the distinct functional role of iT_{reg} cells might be to control inflammation at mucosal surfaces, which are sites of preponderant exposure to non-self substances. This notion is consistent with data indicating that tT_{reg} cells arise from a subset of thymocytes, which exhibit TCR with an increased affinity for self-antigens yet insufficient for negative selection^{10,21}, whereas iT_{reg} cells are efficiently generated

upon TCR engagement with a high affinity cognate ligand under subimmunogenic conditions^{22,23}.

The absence of iT_{reg} cell induction in response to oral antigen in CNS1⁻ mice suggested that the immune balance in the gastrointestinal tract might be impaired owing to deficiency in gut antigen-specific iT_{reg} cells. Indeed, while IL-17 and IFN- γ production by CD4⁺ T cells was unaffected by iT_{reg} deficiency in CNS1⁻ mice (Supplementary Fig. 11), we observed markedly augmented production of the T_H2 cytokines, IL-4, IL-5 and IL-13, by CD4⁺ T cells, especially in the mesenteric lymph nodes, Peyer's patches and intestinal lamina propria (Fig. 2a and Supplementary Fig. 12). Furthermore, the vast majority of CD4⁺ T cells in the lamina propria of CNS1⁻ mice expressed high amounts of Gata3, a key T_H2 differentiation factor. Increases in Gata3⁺CD4⁺ T cells were observed not only in gastrointestinal tract tissues in CNS1⁻ mice but also in other lymphoid tissues, albeit to a lesser extent (Fig. 2b and Supplementary Fig. 12). Consistent with the sharply augmented T_H2 responses at mucosal sites, CNS1⁻ mice exhibited increased frequencies of germinal centre B cells (Fas⁺GL7⁺) in the Peyer's patches, but not in the spleen or peripheral lymph nodes (Supplementary Fig. 13), and spontaneous increases in serum levels of IgE and IgA, but not in other Ig isotypes (Fig. 2c, and data not shown).

The dysregulated T_H2 responses were associated with a decreased body weight (Fig. 3a and Supplementary Fig. 2) and distinct highly penetrant pathology throughout the gastrointestinal tract (Fig. 3b and Supplementary Fig. 14): all CNS1⁻ mice (12/12) and no CNS1^{WT} control littermates (0/6) were affected by gastritis and plasmacytic enteritis characterized by increased frequencies of plasma cells in the intestinal lamina propria and other associated lesions such as crypt abscesses. Accordingly, serum antibodies in CNS1⁻ mice exhibited reactivity against antigens of the small and large intestine, pancreas and chow (Supplementary Fig. 13). Notably, the pathology observed in the gastrointestinal tissue of CNS1⁻ mice was markedly diminished upon B-cell depletion, but was not ameliorated by administration of IL-4 neutralizing antibody (Supplementary Fig. 15). The inflammatory

features and lesions observed in CNS1⁻ mice were consistent with allergic T_H2-type intestinal disease (Fig. 3).

One possible explanation for the pronounced T_H2 responses and intestinal pathology associated with iT_{reg} cell deficiency is simply a numerical decrease in T_{reg} cells. However, we consider this possibility unlikely, because graded depletion of Foxp3⁺ T_{reg} cells in Foxp3^{DTR} mice upon administration of titrated amounts of diphtheria toxin resulting in T_{reg} frequencies similar to those observed in CNS1⁻ mice revealed augmented T_H1 and T_H17, but not T_H2, responses²⁴. Alternatively, certain qualitative features of iT_{reg} cells could allow them to efficiently limit T_H2 inflammation in the gut. Recent studies suggested that some of the transcriptional regulators involved in a particular type of effector T-cell response facilitate the ability of T_{reg} cells to suppress those responses^{25–27}. Thus, we explored the expression of T_H2-associated transcription factor Gata3 in T_{reg} cells in CNS1⁻ and CNS1^{WT} mice. In contrast to a sharp increase in Gata3 expression in effector T cells (Fig. 2b and Supplementary Fig. 12), we found its expression markedly diminished in T_{reg} cells in CNS1⁻ mice (Fig. 3c and Supplementary Fig. 12). Notably, ablation of a conditional Gata3 allele in T_{reg} cells leads to T_{reg} cell dysfunction^{28,29} and marked augmentation of T_H2 cytokine production by CD4⁺ T cells (D. Rudra, R.E.N. and A.Y.R., manuscript in preparation). We hypothesized that increased Gata3 expression in iT_{reg} cells reflects their activation state upon TCR ligation by high affinity ligands in the gut rather than an intrinsic feature of iT_{reg} cells. In support of this idea, we found that both CNS1⁻ and control T_{reg} cells stimulated *in vitro* through the TCR and IL-2 receptor exhibited similarly robust Gata3 induction (Supplementary Fig. 12). Thus, we suggest that increased Gata3 expression in iT_{reg} cells, a likely consequence of their generation in response to high affinity TCR ligands present in the gut, endows these cells with the capacity to efficiently control spontaneous mucosal T_H2 inflammation.

Certain commensal bacteria increase the frequencies of T_{reg} cells in the gut and provide antigens recognized by a considerable proportion of iT_{reg} TCR^{1,16}. In addition to TCR ligands the gut microbial community

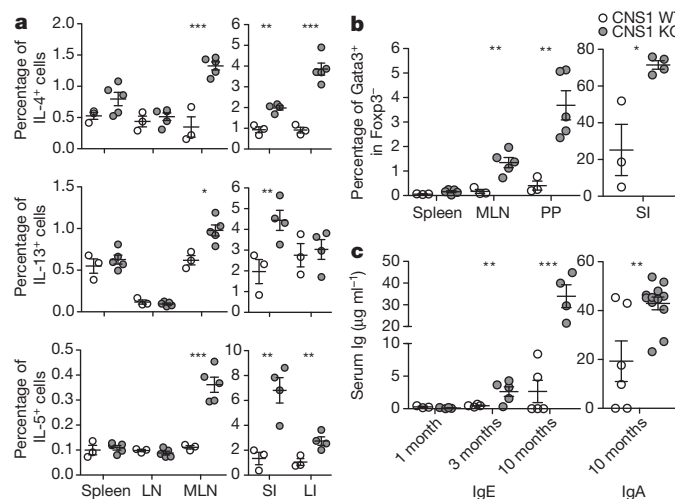


Figure 2 | Paucity of iT_{reg} cells results in T_H2 inflammation in the gastrointestinal tract. **a**, Percentage of CD4⁺ cells producing IL-4 (top), IL-13 (middle) and IL-5 (bottom) in 3-month-old mice. Left, spleen, peripheral lymph nodes (LN) and mesenteric lymph nodes (MLN); right, lamina propria of small and large intestine (SI and LI, respectively). **b**, Percentage of Foxp3⁺CD4⁺ cells that were Gata3⁺ in 3-month-old mice (PP, Peyer's patches). **c**, Concentration of IgE and IgA in serum, determined by enzyme linked immunosorbent assay (ELISA) at 1, 3 and 10 months. All data are representative of three or more independent experiments with ≥ 3 mice per group. Error bars, s.d.; * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, as calculated by Student's *t*-test.

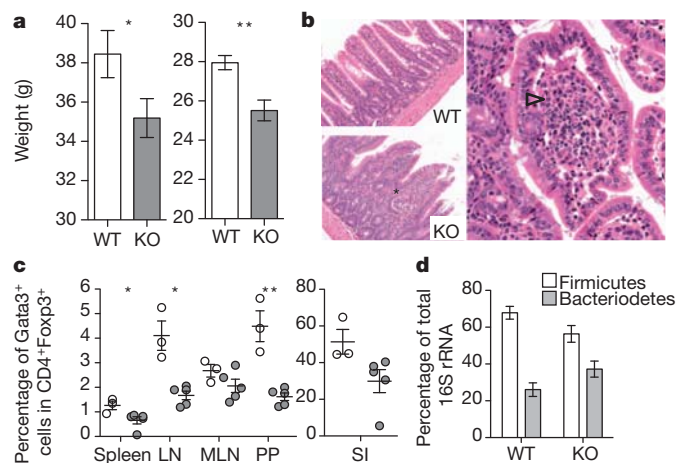


Figure 3 | iT_{reg} cell deficiency leads to T_H2 type gastrointestinal pathology and altered microbial communities. **a**, Body weights of 9–12-month-old individually housed CNS1⁻ (KO) and littermate control (WT) mice ($n \geq 12$). **b**, Plasmacytic enteritis (arrowhead) in CNS1-deficient mice revealed by haematoxylin and eosin staining of small intestine from 9–12-month-old CNS1⁻ (bottom and right) and littermate control mice (top). An early crypt abscess is indicated (asterisk). Data are representative of ≥ 20 mice analysed. **c**, Percentage of Foxp3⁺CD4⁺ cells expressing Gata3⁺ in 3-month-old mice. **d**, Percentage of total 16S rRNA gene sequences of the Firmicutes and Bacteroidetes phyla in stool from individually housed CNS1⁻ ($n = 9$) and WT ($n = 6$) littermate mice. All data are representative of three or more independent experiments with ≥ 3 mice per group. Error bars show s.d. (**a**, **c**) or s.e.m. (**d**). * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, as calculated by Student's *t*-test. Scale bars, 150 μm.

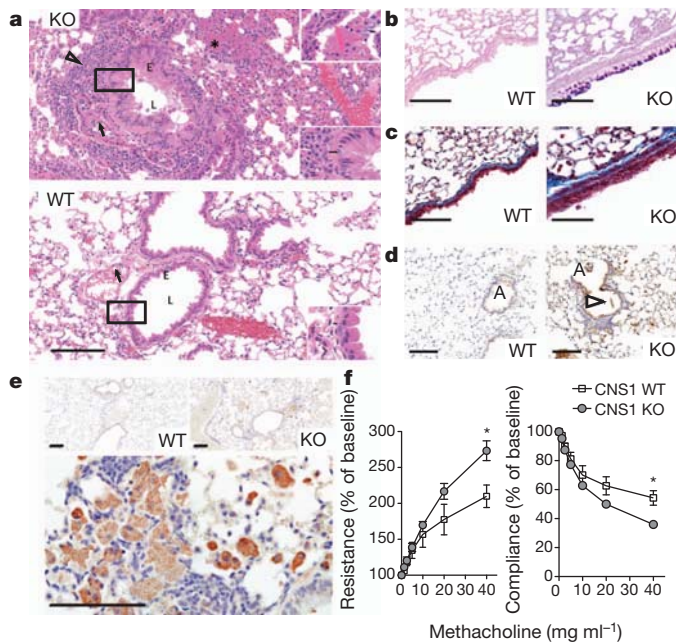


Figure 4 | Unprovoked asthma-like airway pathology in CNS1-deficient mice. **a**, Representative haematoxylin and eosin-stained lung sections from CNS1^{-/-} (top) and WT (bottom) mice. The CNS1^{-/-} lung has marked peribronchiolar inflammation (arrowhead). The reduced lumen (L) contains mucus produced by the hyperplastic respiratory epithelium (E). Arrows indicate reactive (top) and normal (bottom) endothelium. Bottom right hand corner insets are higher magnification of boxed regions and bar indicates smooth muscle thickness. Top right inset (KO) demonstrates eosinophilic crystals. Asterisk marks acidophilic macrophages. **b**, Periodic acid Schiff with Alcian Blue staining highlighting mucus-producing goblet cells (dark blue-purple) **c**, Trichrome staining illustrating lung fibrosis (blue staining). **d**, Arginase-1 staining of lungs from CNS1^{-/-} and WT mice. A indicates airway; an acidophilic crystal is marked by the arrowhead. **e**, Chitinase 3-like 3 (Chi3l3) staining of lungs from CNS1^{-/-} and WT mice at 10× magnification (top) and 60× magnification of lungs from CNS1^{-/-} mice demonstrating robust Chi3l3 expression within acidophilic macrophages (bottom). **f**, Lung resistance (left) and compliance (right) of CNS1^{-/-} and WT littermate control mice after exposure to methacholine. Data representative of two independent experiments with ≥4 mice per group. Error bars, s.d.; **P* < 0.05, ***P* < 0.01, ****P* < 0.001, as calculated by Students' *t*-test. Scale bars, 100 μm.

also contributes to the local cytokine environment, which facilitates iT_{reg} cell differentiation and maintenance in the gut¹. These observations raise a question as to whether iT_{reg} cells, in turn, influence composition of the commensal microbiota. To address this question, we sequenced 16S ribosomal RNA coding genes from bacterial contents of stool samples isolated from CNS1^{-/-} and CNS1^{WT} littermates, which were housed individually for 5 weeks after weaning. Phylogenetic analysis revealed distinct gut microbial communities in CNS1^{-/-} mice, with statistically significant enrichment of the candidate phylum TM7 and the genus *Bacteroidetes Alistipes* (Supplementary Fig. 16), and an overall decrease in the ratio of Firmicutes to Bacteroidetes (2.60 in wild-type and 1.51 in knockout) (Fig. 3d). Interestingly, an opposite trend in the Firmicutes/Bacteroidetes ratio was correlated with obesity³⁰, suggesting the possibility that alterations in energy harvest and metabolism (caused by inflammation or microbe-dependent effects on energy balance) could account for the decreased weight observed in iT_{reg} cell deficient mice. Thus, iT_{reg} cells help maintain a 'normal' microbial community in the gut, probably through exerting control over T_H2 mucosal inflammation.

These observations raised the question of whether the altered microbiota, rather than iT_{reg} deficiency, was the direct cause of observed T_H2 inflammation. To equalize gut microbiota, CNS1^{-/-} and littermate controls were treated with antibiotics (metronidazole and ciprofloxacin) for 4 weeks. Despite indistinguishable microbial communities,

antibiotic treatment did not lead to a decrease in Gata3 expression or Th2 cytokine production by effector T cells in CNS1^{-/-} mice, and characteristic histopathologic features were maintained (Supplementary Fig. 17). Furthermore, iT_{reg} cell sufficient germ-free mice colonized with CNS1^{-/-} or control microbiota exhibited a similar spectrum of T_H1, T_H2 and T_H17 cytokine production and eventual normalization of microbiota (Supplementary Fig. 18 and data not shown). These results suggest that iT_{reg} deficiency results in immune dysregulation and T_H2 inflammation in the gut with subsequent perturbation of the microbial community.

According to the notion of specialized iT_{reg} cell function in suppression of T_H2 responses at mucosal sites, one would expect to observe T_H2-type pathology in the lungs of CNS1^{-/-} mice, despite an only modest ~20–25% decrease in numbers of T_{reg} cells in this tissue compared to littermate controls (Fig. 1c). Indeed, we discovered that CNS1^{-/-} mice suffer from spontaneous T_H2-type airway inflammation (Fig. 4 and Supplementary Fig. 19). The lungs of CNS1^{-/-} mice were characterized by increased infiltration by lymphocytes, plasma cells and macrophages, and by moderate neutrophil infiltration (Fig. 4). The consistent features of the chronic inflammatory airway disease observed in CNS1^{-/-} mice include lymphocytic infiltration, narrowed airway lumen (Fig. 4a), increased goblet cells and mucus production (Fig. 4a and b), smooth muscle hyperplasia, and fibrosis (Fig. 4c). Notably, 9/12 CNS1^{-/-} and 0/6 CNS1^{WT} mice developed acidophilic macrophage pneumonia (AMP) with characteristic increases in acidophilic macrophages and both intracellular and extracellular chitinase 3-like 3 crystals (Chi3l3, formerly Ym1), analogous to Charcott-Lyden crystals found in asthmatic patients (Fig. 4a and e). In addition, the prominent presence of alternatively activated macrophages in the lungs of CNS1^{-/-} mice was confirmed by morphology and expression of arginase 1 in addition to Chi3l3 (Fig. 4d and Supplementary Fig. 20). Furthermore, both young (6–8 week old) and aged (20 week old) CNS1^{-/-} mice exhibited airway hyper-responsiveness accompanied by AMP, perivascular, peribronchiolar and intramucosal inflammation, bronchial epithelial hyperplasia, and airway narrowing (Fig. 4f and Supplementary Fig. 21). These spontaneous lesions are especially striking considering the T_H2-resistant, T_H1-prone C57BL/6 genetic background of CNS1^{-/-} mice. The lung pathology in CNS1^{-/-} mice reflects the hallmark features of chronic allergic inflammation and asthma.

Our results demonstrate that T_{reg} cells of thymic and extrathymic origin have distinct mechanistic requirements for differentiation and exert specialized functions in immune homeostasis. The restriction of lesions to mucosal tissues in iT_{reg} deficient mice implies that under steady state conditions T_{reg} cells generated in the thymus are largely sufficient for control of most immune responses to self-antigens. These findings suggest that in normal animals, T_{reg} cells generated extrathymically in a CNS1-dependent manner play a non-redundant role in control of mucosal allergic Th2 inflammation and asthma.

METHODS SUMMARY

The generation of the following mouse strains has been previously described^{5,20}. CNS1^{-/-} (*Foxp3*^{ΔCNS1}), *Foxp3*^{GFP} and *Foxp3*^{Cre-GFP-ERT2} R26Y. *Rag1*^{-/-} mice were purchased from The Jackson Laboratory, and CD45.1 B6 and *Tcrb*/*Tcrd*^{-/-} mice, along with above strains were maintained in the Sloan Kettering Institute Research Laboratories animal facility in accordance with institutional regulations. Tissues for histologic analysis were fixed in 10% phosphate-buffered formalin and processed routinely for staining. *In vitro* induction assays were performed with 5 × 10⁴ *Foxp3*^{-GFP} CD4⁺ T cells and 5 μg ml⁻¹ of anti-CD3 and anti-CD28 antibody, 100 U ml⁻¹ IL-2, in 96-well, flat-bottom plates. For *in vitro* and transfer experiments, CD4⁺ T cells were pre-enriched using mouse CD4 Dynabeads (L3T4, Invitrogen) and FACS sorted on an LSR-II (BD Biosciences). Intracellular staining for IL-4 used Cytofix/Cytoperm (BD Biosciences), and staining for other cytokines, *Foxp3* and *Gata3*, used the *Foxp3* staining kit (eBiosciences). For measurement of AHR, mice were anaesthetized with pentobarbital and AHR was assessed by invasive measurement of airway resistance using modified version of a described method (Buxco Electronics). 16S rRNA sequencing was performed on a 454 GS

FLX Titanium pyrosequencing platform following the Roche 454 recommended procedures.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 6 July; accepted 6 December 2011.

Published online 8 February 2012.

1. Maloy, K. J. & Powrie, F. Intestinal homeostasis and its breakdown in inflammatory bowel disease. *Nature* **474**, 298–306 (2011).
2. Sakaguchi, S., Yamaguchi, T., Nomura, T. & Ono, M. Regulatory T cells and immune tolerance. *Cell* **133**, 775–787 (2008).
3. Chen, W. *et al.* Conversion of peripheral CD4⁺CD25[−] naive T cells to CD4⁺CD25⁺ regulatory T cells by TGF- β induction of transcription factor Foxp3. *J. Exp. Med.* **198**, 1875–1886 (2003).
4. Tone, Y. *et al.* Smad3 and NFAT cooperate to induce Foxp3 expression through its enhancer. *Nature Immunol.* **9**, 194–202 (2008).
5. Zheng, Y. *et al.* Role of conserved non-coding DNA elements in the *Foxp3* gene in regulatory T-cell fate. *Nature* **463**, 808–812 (2010).
6. Hill, J. A. *et al.* Retinoic acid enhances Foxp3 induction indirectly by relieving inhibition from CD4⁺CD44^{hi} cells. *Immunity* **29**, 758–770 (2008).
7. Nolting, J. *et al.* Retinoic acid can enhance conversion of naive into regulatory T cells independently of secreted cytokines. *J. Exp. Med.* **206**, 2131–2139 (2009).
8. Xu, L. *et al.* Positive and negative transcriptional regulation of the *Foxp3* gene is mediated by access and binding of the Smad3 protein to enhancer I. *Immunity* **33**, 313–325 (2010).
9. Liu, Y. *et al.* A critical function for TGF- β signaling in the development of natural CD4⁺CD25⁺Foxp3⁺ regulatory T cells. *Nature Immunol.* **9**, 632–640 (2008).
10. Ouyang, W., Beckett, O., Ma, Q. & Li, M. O. Transforming growth factor- β signaling curbs thymic negative selection promoting regulatory T cell development. *Immunity* **32**, 642–653 (2010).
11. Curotto de Lafaille, M. A. & Lafaille, J. J. Natural and adaptive foxp3⁺ regulatory T cells: more of the same or a division of labor? *Immunity* **30**, 626–635 (2009).
12. Mucida, D. *et al.* Oral tolerance in the absence of naturally occurring Tregs. *J. Clin. Invest.* **115**, 1923–1933 (2005).
13. Bautista, J. L. *et al.* Intracellular competition limits the fate determination of regulatory T cells in the thymus. *Nature Immunol.* **10**, 610–617 (2009).
14. Leung, M. W., Shen, S. & Lafaille, J. J. TCR-dependent differentiation of thymic Foxp3⁺ cells is limited to small clonal sizes. *J. Exp. Med.* **206**, 2121–2130 (2009).
15. Moran, A. E. *et al.* T cell receptor signal strength in Treg and iNKT cell development demonstrated by a novel fluorescent reporter mouse. *J. Exp. Med.* **208**, 1279–1289 (2011).
16. Lathrop, S. K. *et al.* Peripheral education of the immune system by colonic commensal microbiota. *Nature* **478**, 250–254 (2011).
17. Komatsu, N. *et al.* Heterogeneity of natural Foxp3⁺ T cells: a committed regulatory T-cell lineage and an uncommitted minor population retaining plasticity. *Proc. Natl Acad. Sci. USA* **106**, 1903–1908 (2009).
18. Hori, S. Regulatory T cell plasticity: beyond the controversies. *Trends Immunol.* **32**, 295–300 (2011).
19. Zhou, X. *et al.* Instability of the transcription factor Foxp3 leads to the generation of pathogenic memory T cells *in vivo*. *Nature Immunol.* **10**, 1000–1007 (2009).
20. Rubtsov, Y. P. *et al.* Stability of the regulatory T cell lineage *in vivo*. *Science* **329**, 1667–1671 (2010).
21. Josefowicz, S. Z. & Rudensky, A. Control of regulatory T cell lineage commitment and maintenance. *Immunity* **30**, 616–625 (2009).
22. Gottschalk, R. A., Corse, E. & Allison, J. P. TCR ligand density and affinity determine peripheral induction of Foxp3 *in vivo*. *J. Exp. Med.* **207**, 1701–1711 (2010).
23. Kretschmer, K. *et al.* Inducing and expanding regulatory T cell populations by foreign antigen. *Nature Immunol.* **6**, 1219–1227 (2005).
24. Tian, L. *et al.* Foxp3 regulatory T cells exert asymmetric control over murine helper responses by inducing Th2 cell apoptosis. *Blood* **118**, 1845–1853 (2011).
25. Chaudhry, A. *et al.* CD4⁺ regulatory T cells control T_H17 responses in a Stat3-dependent manner. *Science* **326**, 986–991 (2009).
26. Koch, M. A. *et al.* The transcription factor T-bet controls regulatory T cell homeostasis and function during type 1 inflammation. *Nature Immunol.* **10**, 595–602 (2009).
27. Zheng, Y. *et al.* Regulatory T-cell suppressor program co-opts transcription factor IRF4 to control T_H2 responses. *Nature* **458**, 351–356 (2009).
28. Wang, Y., Su, M. A. & Wan, Y. Y. An essential role of the transcription factor GATA-3 for the function of regulatory T cells. *Immunity* **35**, 337–348 (2011).
29. Wohlfert, E. A. *et al.* GATA3 controls Foxp3⁺ regulatory T cell fate during inflammation in mice. *J. Clin. Invest.* **121**, 4503–4515 (2011).
30. Turnbaugh, P. J. *et al.* An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**, 1027–1031 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank T. Tedder for depleting CD20 antibody, R. Tudor for assistance interpreting lung pathology, P. DeRoos for assistance with Ig ELISA assays, B. Johnson for immunohistochemical expertise, Y. Chen for assistance with airway measurements, and E. Pamer, L. Lipuma, A. Gbourne and R. Khanin for help with analysis of intestinal microbiota. This work was supported by NIH MSTP grant GM07739 and NINDS grant 1F31NS073203-01 (R.E.N.), Strategic Young Researcher Overseas Visits Program for Accelerating Brain Circulation from Department of Microbiology and Immunology, Keio University School of Medicine (T.C.) and NIH grant R37 AI034206 (A.Y.R.). A.Y.R. is an investigator with the Howard Hughes Medical Institute.

Author Contributions S.Z.J., R.E.N. and H.Y.K. performed experiments and analysed data, with assistance from T.C. for tissue Ig ELISA experiments, and P.T. for immunohistochemistry and histopathology analysis. D.T.U., S.Z.J., R.E.N., H.Y.K. and A.Y.R. designed and interpreted AHR experiments. Y.Z. generated CNS1[−] mice. S.Z.J., R.E.N. and A.Y.R. designed experiments and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.Y.R. (rudenska@mskcc.org).

METHODS

Mice. The generation of the following mouse strains has been previously described^{5,20}: *CNS1*[−] (*Foxp3*^{ΔCNS1}), *Foxp3*^{GFP} and *Foxp3*^{Cre-GFP-ERT2} R26Y. *Rag1*[−] mice were purchased from The Jackson Laboratory, and CD45.1 B6 and *Tcrb/Tcrd*^{−/−} mice, along with above strains were maintained in the Sloan Kettering Institute Research Laboratories animal facility in accordance with institutional regulations. Mice were killed by CO₂ asphyxiation. EAE was induced and scored as previously described³¹. For antibiotic treatment, *CNS1*-deficient and sufficient mice were treated with 1 g l^{−1} metronidazole (Sigma-Aldrich) and 0.2 g l^{−1} ciprofloxacin (ENZO Life Sciences International) dissolved in drinking water for 4 weeks. Mouse anti-CD20⁸ (MB20-11, provided by T. Tedder) and anti-IL-4 (11b.11, NCI-Frederick) were administered weekly as intraperitoneal injections of 50 μg or 5 μg, respectively, for 3 weeks.

Cell isolation, transfer and FACS staining. For *in vitro* and *in vivo* transfer experiments, CD4⁺ T cells were pre-enriched using mouse CD4 Dynabeads (L3T4, Invitrogen) and FACS sorted on an LSR-II (BD Biosciences). Intracellular staining for IL-4 used Cytofix/Cytoperm following treatment with Golgi-Stop (BD Biosciences), and staining for other cytokines (following treatment with Golgi-Plug, BD Biosciences) and Foxp3 and Gata3 used the Foxp3 staining kit (eBiosciences).

In vitro assays. *In vitro* induction assays were performed with 5 × 10⁴ Foxp3[−] CD4⁺ T cells and 5 μg ml^{−1} of anti-CD3 and anti-CD28 antibody, 100 U ml^{−1} IL-2, in 96-well, flat-bottom plates. For *in vitro* suppression assays, 4 × 10⁴ CD4⁺ Foxp3[−] CD62L^{high} naive T cells FACS purified from WT mice were cultured with graded numbers of CD4⁺ Foxp3⁺ T_{reg} cells FACS purified from *Foxp3*^{ΔCNS1} or *Foxp3*^{GFP} mice in the presence of 10⁵ irradiated T cell-depleted splenocytes and 1 μg ml^{−1} anti-CD3 antibody in a 96-well round-bottom plate for 80 h. Cell proliferation was assessed by [³H]thymidine incorporation during the final 8 h of culture.

Histology and immunohistochemistry. Necropsies were performed, and sections of pancreas, stomach, heart, lungs, kidney, external ear and haired skin were fixed in 10% phosphate-buffered formalin. Tissues were processed routinely for staining with haematoxylin and eosin, periodic acid Schiff with Alcian blue or Masson Trichrome if indicated. Slides were examined by an American Board of Veterinary Practitioners-certified veterinary pathologist blinded to genotypes. Morphological diagnoses were applied for all tissues. Immunohistochemical staining was performed by the University of Washington Histology and Imaging Core using standard protocols with a Leica Bond Automated Immunostainer. Primary antibodies: goat anti-mouse chitinase 3-like 3/ECF-L (YM1) (R&D systems, cat. no. AF2446, lot no. UNU01), 0.2 μg ml^{−1}; rabbit polyclonal anti iNOS/NOS II, NT (Millipore, cat. no. 06-573), 1 μg ml^{−1}; rabbit polyclonal anti arginase 1 (H-52) (Santa Cruz, cat. no. sc-20150, lot no. K0807), 0.2 μg ml^{−1}. Isotype controls were used at the same concentration as the primary antibody with all antibodies run with Leica Bond reagents and Bond Polymer Refine (DAB) detection with haematoxylin counter stain.

Histology inflammation scoring. 0, None; 1, focal or multifocal mild perivascular accumulations with minimal extension into surrounding adventia or parenchyma; 2, multifocal mild or focal moderate perivascular accumulations with mild extension into surrounding parenchyma or mild to moderate parenchymal accumulations; 3, grade 2 plus mild inflammation-associated parenchymal lesions such as loss or degeneration of cells; 4, grade 2 plus moderate to severe inflammation-associated parenchymal lesions. Inflammation in the gastrointestinal tract was scored as described previously³².

Airway hyperresponsiveness measurements. For measurement of AHR, mice were anaesthetized with pentobarbital (7.5–10 mg per mouse) and AHR was assessed by invasive measurement of airway resistance using modified version of a described method (Buxco Electronics). Mice were ventilated at a tidal volume of 0.2 ml with the use of a ventilator (Harvard Apparatus) and frequency was set around 150 Hz. Baseline pulmonary mechanics and responses to ventilated saline (0.9% NaCl) were measured, and lung resistance (*R*_L) was measured in response to increasing doses (0.125–40 mg ml^{−1}) of acetyl-β-methylcholine chloride (methacholine; MCh) (Sigma-Aldrich). The three values of *R*_L obtained after each dose of methacholine were averaged to obtain the final values for each dose. Results are expressed as percentage of increase of saline-baseline. Following measurement of AHR, mouse tracheas were cannulated and the lungs were lavaged twice with 1 ml of PBS 2% FCS and the fluids were pooled. Cells in the lavage fluid were counted using a haemocytometer, and BAL cell differential counts were determined on slide preparations stained with DiffQuik. At least 200 cells were differentiated on stained slides by light microscopy using conventional morphological criteria. For some experiments, BAL for each mouse or grouped BAL was stained and analysed by flow cytometry.

Stool sample collection. Fresh stool samples were induced directly into sterile collection tubes from live *CNS1*[−] and control mice and snap frozen before preparation of material for sequencing (see below).

DNA extraction. DNA extraction was performed on each fecal specimen using phenol-chloroform extraction with mechanical disruption based on a previously described protocol³³. Briefly, an aliquot (~500 mg) of each sample was suspended in a solution containing 500 μl of extraction buffer (200 mM Tris, pH 8.0; 200 mM NaCl; and 20 mM EDTA), 210 μl of 20% SDS, 500 μl of phenol/chloroform/isoamyl alcohol (25:24:1), and 500 μl of 0.1-mm-diameter zirconia/silica beads (BioSpec Products). Microbial cells were lysed by mechanical disruption with a bead beater (BioSpec Products) for 2 min, after which two rounds of phenol/chloroform/isoamyl alcohol extraction were performed. DNA was precipitated with ethanol and resuspended in 50 μl of nuclease-free water. DNA was subjected to additional purification with the QIAamp DNA Mini Kit (Qiagen).

PCR amplification and sequencing. For each sample, three replicate 25 μl PCR amplifications were performed, each containing 5 ng of purified DNA, 0.2 mM dNTPs, 1.5 mM MgCl₂, 1.25 U Platinum Taq DNA polymerase, 2.5 μl of 10× PCR buffer, and 0.2 μM each of broad-range bacterial forward and reverse primers as described previously³⁴, flanking the V1–V3 variable region. The primers were modified to include adaptor sequences required for 454 sequencing, with the addition of a unique 6–8 base barcode in the reverse primer. The forward primer (5′-CCTATCCCTGTGTGCCTTGGCAGTCTCAGAGTTTGATCCTGGCTCAG-3′) consisted of the 454 Lib-L primer B (underlined) and the broad-range universal bacterial primer 8F (italics); the reverse primer (5′-CCATCTCATCCC TGCGTGTCTCCGACTCAGNNNNNNNATTACGCGGCTGCTGG-3′) consisted of the 454 Lib-L primer A, barcode (NNNNNNN), and the broad-range primer 534R (italics). The cycling conditions were: 94 °C for 3 min, then 25 cycles of 94 °C for 30 s, 56 °C for 30 s, and 72 °C for 1 min. The three replicate PCR products were pooled and subsequently purified using the Qiaquick PCR Purification Kit (Qiagen). The purified PCR products were sequenced unidirectionally on a 454 GS FLX Titanium pyrosequencing platform following the Roche 454 recommended procedures.

Sequence processing and analysis. Sequences were converted to standard FASTA format using Vendor 454 software. Sequences shorter than 200 base pairs (bp), containing undetermined bases or homopolymer stretches longer than 8 bp, or failing to align with the V1–V3 region were excluded from the analysis. Using the 454 base quality scores, which range from 0 to 40 (0 being an ambiguous base), sequences were trimmed using a sliding-window technique, such that the minimum average quality score over a window of 50 bases never dropped below 35. Sequences were trimmed from the 3′-end until this criterion was met. Sequences were aligned to the V1–V3 region of the 16S gene, using as template the SILVA reference alignment³⁵ and the Needleman-Wunsch algorithm with default scoring options. Potentially chimaeric sequences were removed using the chimera uclime program³⁶. Sequences were grouped into operational taxonomic units (OTUs) using the average neighbour algorithm. Sequences with distance-based similarity of 97% or greater were assigned to the same OTU. For each fecal sample, OTU-based microbial diversity was estimated by calculating the Shannon diversity index³⁷. Phylogenetic classification to genus level was performed for each sequence, using the Bayesian classifier algorithm described by Wang and colleagues, using a database of known 16S sequences generated by the Ribosomal Database Project (RDP)³⁸. For each experiment, data were analysed on each taxon level individually. The count data was rescaled using DESeq R package³⁹. Bacteria with less than 10 mean count in both conditions were removed from further analysis and bacteria with statistically significant differences between two conditions (for example, WT and KO), were determined using binomial test (from DESeq package). Bacteria with fold-change greater than two and FDR = 0.05 were declared significant.

- Stromnes, I. M. & Gorman, J. M. Active induction of experimental allergic encephalomyelitis. *Nature Protocols* **1**, 1810–1819 (2006).
- Burich, A. et al. Helicobacter-induced inflammatory bowel disease in IL-10- and T cell-deficient mice. *Am. J. Physiol. Gastrointest. Liver Physiol.* **281**, G764–G778 (2001).
- Ubeda, C. et al. Vancomycin-resistant Enterococcus domination of intestinal microbiota is enabled by antibiotic treatment in mice and precedes bloodstream invasion in humans. *J. Clin. Invest.* **120**, 4332–4341 (2010).
- Nossa, C. W. et al. Design of 16S rRNA gene primers for 454 pyrosequencing of the human foregut microbiome. *World J. Gastroenterol.* **16**, 4135–4144 (2010).
- Schloss, P. D. et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **75**, 7537–7541 (2009).
- Edgar, R. C. et al. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27**, 2194–2200 (2011).
- Magurran, A. E. *Measuring Biological Diversity* (Blackwell, 2004).
- Wang, Q. et al. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* **73**, 5261–5267 (2007).
- Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010).

Cancer exome analysis reveals a T-cell-dependent mechanism of cancer immunoediting

Hirokazu Matsushita^{1†*}, Matthew D. Vesely^{1*}, Daniel C. Koboldt², Charles G. Rickert¹, Ravindra Uppaluri³, Vincent J. Magrini^{2,4}, Cora D. Arthur¹, J. Michael White¹, Yee-Shiuan Chen¹, Lauren K. Shea¹, Jasreet Hundal², Michael C. Wendl^{2,4}, Ryan Demeter², Todd Wylie², James P. Allison^{5,6}, Mark J. Smyth^{7,8}, Lloyd J. Old⁹, Elaine R. Mardis^{2,4} & Robert D. Schreiber¹

Cancer immunoediting, the process by which the immune system controls tumour outgrowth and shapes tumour immunogenicity, is comprised of three phases: elimination, equilibrium and escape^{1–5}. Although many immune components that participate in this process are known, its underlying mechanisms remain poorly defined. A central tenet of cancer immunoediting is that T-cell recognition of tumour antigens drives the immunological destruction or sculpting of a developing cancer. However, our current understanding of tumour antigens comes largely from analyses of cancers that develop in immunocompetent hosts and thus may have already been edited. Little is known about the antigens expressed in nascent tumour cells, whether they are sufficient to induce protective antitumour immune responses or whether their expression is modulated by the immune system. Here, using massively parallel sequencing, we characterize expressed mutations in highly immunogenic methylcholanthrene-induced sarcomas derived from immunodeficient *Rag2*^{−/−} mice that phenotypically resemble nascent primary tumour cells^{1,3,5}. Using class I prediction algorithms, we identify mutant spectrin-β2 as a potential rejection antigen of the d42m1 sarcoma and validate this prediction by conventional antigen expression cloning and detection. We also demonstrate that cancer immunoediting of d42m1 occurs via a T-cell-dependent immunoselection process that promotes outgrowth of pre-existing tumour cell clones lacking highly antigenic mutant spectrin-β2 and other potential strong antigens. These results demonstrate that the strong immunogenicity of an unedited tumour can be ascribed to expression of highly antigenic mutant proteins and show that outgrowth of tumour cells that lack these strong antigens via a T-cell-dependent immunoselection process represents one mechanism of cancer immunoediting.

For this study, we chose two representative, highly immunogenic, unedited methylcholanthrene (MCA)-induced sarcoma cell lines, d42m1 and H31m1, derived from immunodeficient *Rag2*^{−/−} mice¹. Both grow progressively when transplanted orthotopically into *Rag2*^{−/−} mice, but are rejected when transplanted into naive wild-type mice (Supplementary Figs 1 and 2). Using a modified form of exome sequencing involving complementary DNA (cDNA) capture by mouse exome probes and Illumina deep sequencing (that is, cDNA capture sequencing or cDNA CapSeq), we identified 3,737 somatic, non-synonymous mutations in d42m1 cells (3,398 missense, 221 nonsense, 2 nonstop and 116 splice site mutations) and 2,677 non-synonymous mutations in H31m1 cells (2,391 missense, 160 nonsense, 3 nonstop and 123 splice site mutations) (Fig. 1a and Supplementary Fig. 3 and Supplementary Table 1). The mutations in each cell line

were largely distinct—d42m1 and H31m1 share only 119 identical missense mutations (Fig. 1b and Supplementary Table 2)—a result that potentially explains the unique antigenicity of each cell line (Supplementary Fig. 4). Although d42m1 and H31m1 display mutations in known cancer genes⁶, the functional effects of these novel mutations remain undefined. Nevertheless, both tumours have cancer-causing mutations in *Kras* (codon 12) and *Trp53* that are frequently observed in human and mouse cancers^{7–9} (Supplementary Table 3). The mutation calls were confirmed by independent Roche/454 pyrosequencing of 22 genes using tumour genomic DNA and by documenting their absence in normal cells from the same mouse that developed the tumour (Supplementary Table 4).

Comparing cDNA CapSeq data of d42m1 and H31m1 cells to human cancer genomes^{10–17} revealed two similarities. First, 46–47% of mutations in d42m1 and H31m1 are C/A or G/T transversions, which represent chemical-carcinogen signatures^{7,13,14} similar to those of lung cancers from smokers (44–46%) but not seen in human cancers induced by other mechanisms (8–16%) (Fig. 1c). Second, the mutation rates of d42m1 and H31m1 are about tenfold higher than those of lung cancers from smokers, but within threefold of hypermutator smoker lung cancers with mutations in DNA repair pathway genes (Fig. 1d). Interestingly, d42m1 and H31m1 also show mutations in DNA repair genes (Supplementary Table 3), although these novel mutations have not been functionally characterized. Thus, mouse MCA-induced sarcomas have qualitative and quantitative genomic similarities to carcinogen-induced human cancers.

When parental d42m1 sarcoma cells were transplanted into naive wild-type mice, approximately 20% of recipients developed escape tumours (Supplementary Fig. 5a, c). Cell lines made from three escape tumours (d42m1-es1, d42m1-es2 and d42m1-es3) formed progressively growing sarcomas when transplanted into naive wild-type recipients (Fig. 2a). In contrast, parental d42m1 tumour cells passaged through *Rag2*^{−/−} mice maintained high immunogenicity (Supplementary Fig. 5b, d). Additional analyses revealed that whereas eight of ten clones of d42m1 were rejected in wild-type mice, two clones (d42m1-T3 and d42m1-T10) grew with kinetics similar to d42m1 escape tumours (Fig. 2a and Supplementary Fig. 6). Thus, the d42m1 cell line consists mostly, but not entirely, of highly immunogenic clones and undergoes immunoediting in wild-type mice. cDNA CapSeq of parental d42m1 cells, clones and escape tumours revealed that all expressed similar numbers of mutations (Supplementary Fig. 7a and Supplementary Table 1) and phylogenetic analysis revealed that all d42m1-derived cells were genomically related to one another but distinct from H31m1 and normal fibroblasts (Supplementary

¹Department of Pathology and Immunology, Washington University School of Medicine, 660 South Euclid Avenue, St Louis, Missouri 63110, USA. ²The Genome Institute, Washington University School of Medicine, 660 South Euclid Avenue, St Louis, Missouri 63110, USA. ³Department of Otolaryngology, Washington University School of Medicine, 660 South Euclid Avenue, St Louis, Missouri 63110, USA. ⁴Department of Genetics, Washington University School of Medicine, 660 South Euclid Avenue, St Louis, Missouri 63110, USA. ⁵Ludwig Center for Cancer Immunotherapy, Department of Immunology, Memorial Sloan-Kettering Cancer Center, New York, New York 10021, USA. ⁶Howard Hughes Medical Institute, Memorial Sloan-Kettering Cancer Center, New York, New York 10021, USA. ⁷Cancer Immunology Program, Peter MacCallum Cancer Centre, East Melbourne, 3002 Victoria, Australia. ⁸Department of Pathology, University of Melbourne, Parkville, 2010 Victoria, Australia. ⁹New York Branch of The Ludwig Institute for Cancer Research at Memorial Sloan-Kettering Cancer Center, New York, New York 10021, USA. [†]Present address: Department of Immunotherapeutics (Medinet), The University of Tokyo Hospital, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8655, Japan.

*These authors contributed equally to this work.

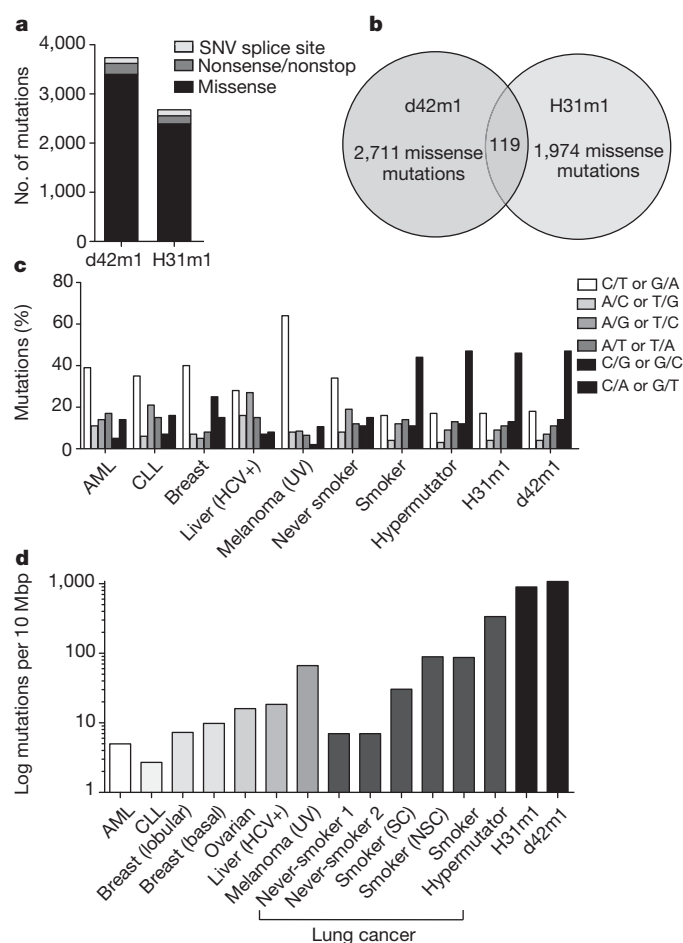


Figure 1 | Unedited MCA-induced sarcomas d42m1 and H31m1 genomically resemble carcinogen-induced human cancers. **a**, Number of non-synonymous mutations in d42m1 and H31m1 tumour cells as detected by cDNA CapSeq. SNV, single nucleotide variant. **b**, Missense mutations compared between d42m1 and H31m1 that had at least 20 \times sequencing coverage. **c**, Spectrum of DNA nucleotide substitutions detected in d42m1 and H31m1 as compared to previously generated data from human cancers including acute myelogenous leukaemia¹⁰ (AML), chronic lymphocytic leukaemia¹⁶ (CLL), breast cancer (breast lobular¹², breast basal¹¹), ovarian cancer (E. R. Mardis *et al.*, manuscript in preparation), liver cancer (hepatitis C virus (HCV)-positive)¹⁵, melanoma (ultraviolet (UV)-induced)¹⁷ and lung cancers (non-small cell (NSC)¹³, small cell (SC)¹⁴, never-smoker, smoker and hypermutator (E. R. Mardis *et al.*, manuscript in preparation). **d**, Mutation rates for d42m1, H31m1 and human cancers described in **c** including tumours from never-smoker 1 (bronchioloalveolar carcinoma) and never-smoker 2 (lung adenocarcinoma).

Fig. 7b). However, regressor clones clustered more closely to parental d42m1 cells whereas progressor clones clustered more closely to cells from escape tumours. Thus, the d42m1 tumour cell line consists of a related, but heterogeneous population of tumour cells.

Tumour-specific mutant proteins presented on mouse or human MHC class I molecules are known to represent one class of tumour-specific antigens for CD8⁺ T cells^{18,19}. Therefore, we used *in silico* analysis²⁰ to assess the theoretical capacities of missense mutations from d42m1-related tumour cells to bind MHC class I proteins. Each d42m1-related cell type expressed many potential high-affinity (half-maximum inhibitory concentration (IC₅₀) < 50 nM; affinity value (1/IC₅₀ \times 100) > 2) epitopes that could bind to H-2D^b or H-2K^b (Fig. 2b). Of these, 39–42 were expressed only in the regressor subset of d42m1-related cells (7–9 for H-2D^b, 30–35 for H-2K^b), including 31 expressed in all regressor cells (Supplementary Table 5). Thus, ~1% of the missense mutations in d42m1 are selectively expressed in rejectable d42m1 clones.

Whereas parental and regressor d42m1 cells stimulated interferon- γ (IFN- γ) release *in vitro* when incubated with a specific CD8⁺ cytotoxic T lymphocyte (CTL) clone (C3) derived from a wild-type mouse that had rejected parental d42m1 tumour cells (Fig. 3a, b), progressor d42m1 clones, cells from escape tumours or unrelated MCA sarcomas did not. This result demonstrated that all regressor d42m1 tumour cells share a mutation that forms the epitope recognized by C3 CTLs. As recognition of d42m1 regressor cells by C3 CTLs is restricted by H-2D^b (Fig. 3c), we postulated that an R913L mutation in spectrin- β 2 produced the most likely target for C3 CTLs because its expression was restricted to d42m1 regressor clones and it formed an epitope that showed high-affinity binding potential to H-2D^b in contrast to the wild-type sequence predicted to bind with low affinity (Fig. 3d and Supplementary Table 5).

To verify the importance of mutant spectrin- β 2 on d42m1 antigenicity, we independently identified the tumour antigen recognized by the C3 CTL clone using a T-cell-based expression cloning approach²¹. After three screening rounds, a single positive cDNA was identified encoding a sequence identical to the R913L spectrin- β 2 mutant (Fig. 3e). Thus, conventional antigen expression cloning identified the same mutation predicted by the genomic sequencing.

Mutation-specific real-time quantitative polymerase chain reaction with reverse transcription (qRT-PCR) revealed the presence of mutant spectrin- β 2 messenger RNA in parental d42m1 tumour cells and regressor d42m1 clones, but not in progressor d42m1 clones or escape tumours (Fig. 3f), nor in normal tissue of the mouse from which the d42m1 tumour was derived (Supplementary Table 4 and Supplementary Fig. 8). Additionally, C3 CTLs discriminated between mutant and wild-type spectrin- β 2 peptide sequences when presented on an unrelated H-2D^b-expressing cell line (Fig. 3g). Whereas the mutant (VAVVNQIAL; underline letter indicates the site of mutation) peptide stimulated C3 CTLs in a dose-dependent manner, the wild-type (VAVVNQIAR) peptide did not, even when added in 1,000-fold excess. Using labelled H-2D^b tetramers generated with mutant peptide, mutant spectrin- β 2-specific CD8⁺ T cells accumulated over time in parental d42m1 tumours developing *in vivo* and draining lymph nodes before tumour rejection (Fig. 4a, b). In contrast, no mutant spectrin- β 2-specific CD8⁺ T cells were detected in progressively growing escape tumours or draining lymph nodes. These data demonstrate that mutant spectrin- β 2 expressed selectively in a high proportion of unedited d42m1 tumour cells evokes a T-cell response in naive wild-type mice that promotes the elimination of antigen-expressing tumour cells.

To test whether expression of mutant spectrin- β 2 was sufficient to drive rejection of d42m1 tumour cells, we enforced expression of either mutant or wild-type spectrin- β 2 in d42m1-es3 cells that lack this mutation (Supplementary Fig. 9a) and followed their growth in wild-type mice. Whereas d42m1-es3 tumour cell clones transduced with either control retrovirus or retrovirus encoding wild-type spectrin- β 2 (WT.1 and WT.3) grew progressively with growth kinetics similar to unmanipulated d42m1-es3 cells, d42m1-es3 clones expressing mutant spectrin- β 2 (mu.6 and mu.14) were rejected in wild-type mice, but not in Rag2^{-/-} mice (Fig. 4c and Supplementary Fig. 9b, c, d). CD8⁺ T cells specific for mutant spectrin- β 2 did not infiltrate d42m1-es3 tumours expressing wild-type spectrin- β 2 (WT.3), but were present in d42m1-es3 tumours expressing mutant spectrin- β 2 (mu.14) that were rejected in wild-type mice (Fig. 4d). Thus, mutant spectrin- β 2 is indeed a major rejection antigen of d42m1 sarcoma cells and d42m1 escape from immune control is the consequence of outgrowth of d42m1 clones that lack expression of dominant rejection antigens.

The possibility that the lack of dominant rejection antigen(s) in a small subset of d42m1 cells was due to epigenetic silencing was ruled out because no spectrin- β 2 mutation was (1) found by sequencing genomic DNA from progressor d42m1 clones or escape tumours (Supplementary Table 4) or (2) expressed in d42m1 progressor clones or escape tumours after treatment with inhibitors of methyltransferases and histone deacetylases (Supplementary Fig. 10). We therefore

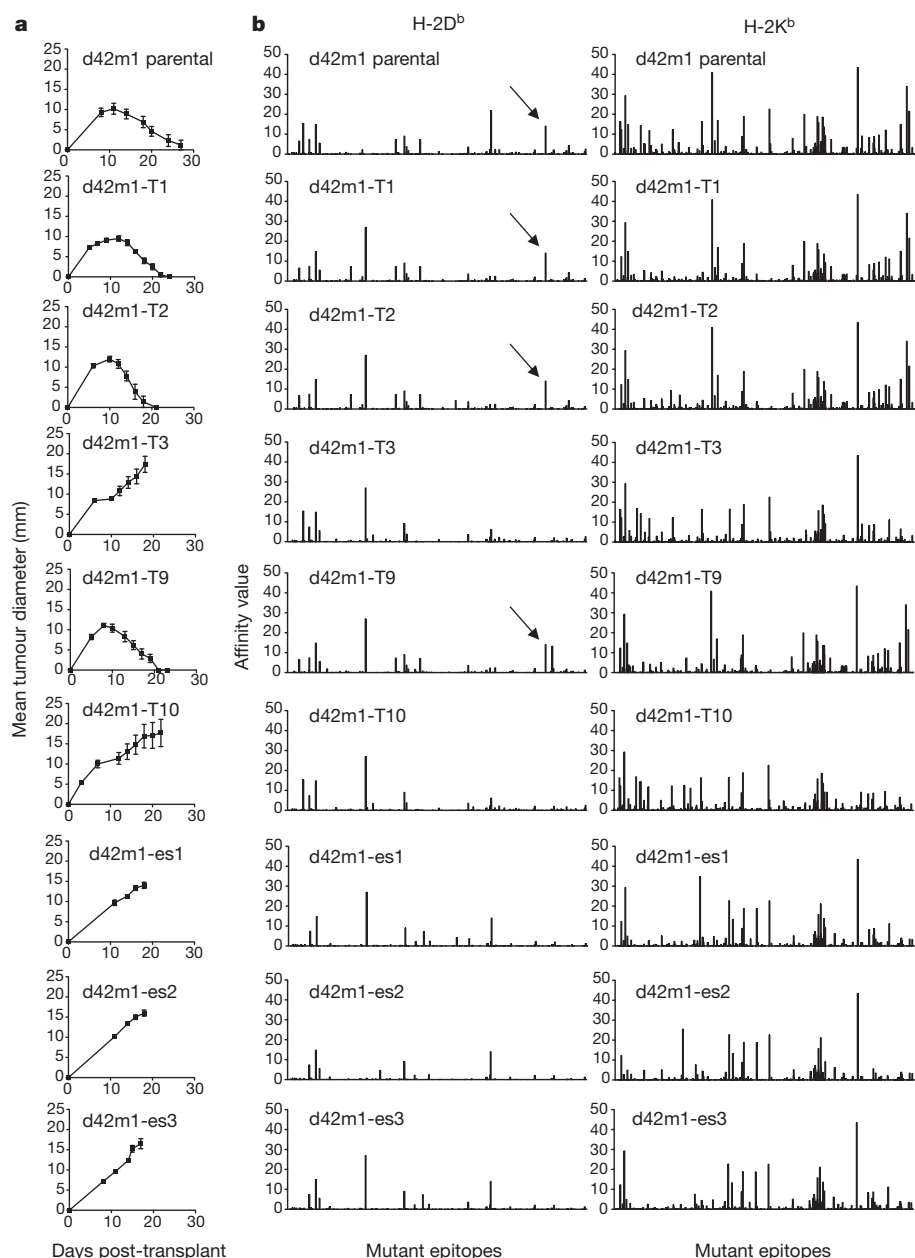


Figure 2 | Affinity value profiles of predicted MHC class I epitopes from tumour-specific mutations. **a**, Growth of d42m1 parental cells, a representative sample of tumour clones, and three escape tumours following transplantation into wild-type mice ($n = 5$, squares). Data are presented as average tumour diameter \pm s.e.m. and are representative of three independent experiments. **b**, Missense mutations for each d42m1-related tumour examined in **a** were analysed for potential MHC class I neoepitopes that bind either H-2D^b or H-2K^b. Predicted epitope binding affinities were ultimately expressed as affinity values ($1/\text{IC}_{50} \times 100$). Arrows indicate H-2D^b epitopes created by the R913L spectrin-β2 mutant.

asked whether T-cell-dependent immunoselection explained the outgrowth of escape tumours. Specifically, we examined the *in vivo* growth behaviour of a tumour cell mixture containing a vast majority of highly immunogenic, mutant spectrin-β2⁺ d42m1-T2 cells and a minority of mutant spectrin-β2⁻ d42m1-T3 progressor cells. To distinguish between the two cell types, we labelled d42m1-T2 with red fluorescent protein (RFP) (modified to eliminate class I epitopes) and d42m1-T3 with green fluorescent protein (GFP) and documented that the labelling did not alter their *in vivo* growth characteristics. We found that we could recapitulate the tumour growth phenotype of parental d42m1 at a ratio of 95% d42m1-T2 cells to 5% d42m1-T3 cells (Fig. 4e). At this ratio, 100% of Rag2^{-/-} mice and wild-type mice depleted of either CD4⁺ or CD8⁺ T cells developed progressively growing tumours (Fig. 4f). In contrast, 5/20 (25%) wild-type mice injected with the tumour cell mixture developed escape tumours, a result that recapitulated the behaviour of parental d42m1. Tumours harvested from Rag2^{-/-} mice were comprised of 84% d42m1-T2 cells and 14% d42m1-T3 cells (Fig. 4h) and expressed mutant spectrin-β2 (Fig. 4g), that is, they resembled the initial 95:5 cell mixture. In contrast, tumours that grew out in wild-type mice consisted of 98%

d42m1-T3 tumour cells and lacked mutant spectrin-β2 (Fig. 4g, h). Thus, d42m1 escape tumours develop as a consequence of T-cell-dependent immunoselection favouring the outgrowth of tumour cells that lack major rejection antigens.

This report shows that the combination of cancer exome sequencing and *in silico* epitope prediction algorithms can identify highly immunogenic, tumour-specific mutational antigens in unedited carcinogen-induced cancers that serve as targets for the elimination phase of cancer immunoediting. To our knowledge, this is the first study to use a genomics approach to experimentally identify a tumour antigen, to specifically identify an antigen from an unedited tumour and to demonstrate that T-cell-dependent immunoselection is a mechanism underlying the outgrowth of tumour cells that lack strong rejection antigens. This mechanism most likely also produces other types of escape tumours, such as those that develop inactivating mutations in antigen presentation genes (for example, those encoding MHC class I proteins), which are frequently observed in clinically apparent human cancers^{22,23}. Developing carcinogen-induced tumours (for example, mouse MCA sarcomas or human smoker lung cancers) may be the preferred targets of cancer immunoediting because they express the

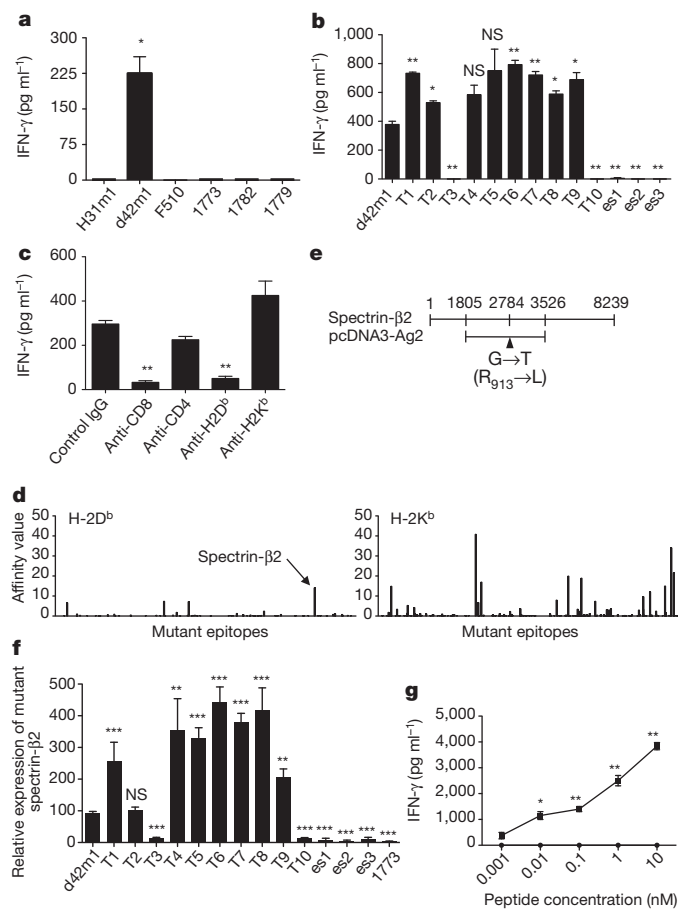


Figure 3 | Identification of mutant spectrin- β 2 as an authentic antigen of an unedited tumour. **a**, **b**, IFN- γ release by C3 CTLs following co-culture with different unedited sarcomas (**a**) or d42m1-related tumours (**b**). **c**, IFN- γ release by C3 CTLs is inhibited by monoclonal antibodies that block CD8 and H-2D^b, but not CD4 or H-2K^b. **d**, MHC class I epitopes predicted to be shared in all of the regressor d42m1 tumours, but not in progressor d42m1 tumours. **e**, Representation of the cDNA clone that stimulated C3 CTLs encoding the spectrin- β 2 R913L mutation. **f**, qRT-PCR for mutant spectrin- β 2 in d42m1-related tumours and 1773. **g**, IFN- γ release by C3 CTLs incubated with COS-D^b cells pulsed with wild-type (circles) or mutant (squares) spectrin- β 2 peptides. Data are representative of three independent experiments. Samples were compared in **b**, **f** to d42m1 using an unpaired, two-tailed Student's *t* test (**P* < 0.05, ***P* < 0.01, ****P* < 0.001; NS, not significant).

greatest number of mutations that might function as neoantigens. However, as ~1% of the mutations in d42m1 are selectively expressed in regressor tumour clones, it is possible that spontaneous tumours arising by other means that harbour as few as 100–200 mutations could still be susceptible to immunological sculpting as they develop. In this regard it is significant that, as documented in a complementary study reported in this issue²⁴, oncogene-induced primary sarcomas engineered to express a strong model antigen can also undergo T-cell-dependent immunoediting, resulting in the outgrowth of tumours that escape immune control. It will be interesting in the future to compare the effects of immunity on the antigenic profiles of oncogene- versus carcinogen-induced tumours.

The immunodominance of mutant spectrin- β 2 in driving tumour rejection in many ways resembles that of certain viral antigens²⁵ and is probably due to the presence in d42m1 of four copies of chromosome 11, each of which carries the spectrin- β 2 gene, thereby producing a highly abundant neoepitope that binds to H-2D^b 750-fold stronger than that of the wild-type sequence. More work is needed to determine which of the other mutations, if any, selectively expressed in d42m1 regressors function as rejection antigens. Immunoepitope analysis of parental

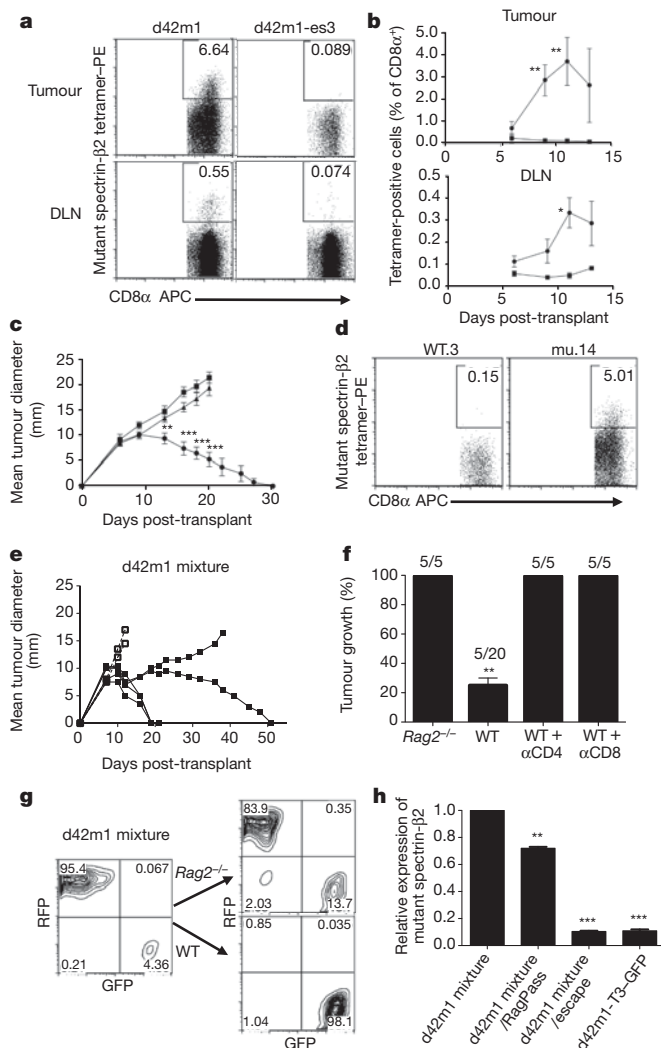


Figure 4 | Mutant spectrin- β 2 is a major rejection antigen of d42m1.

a, Mutant spectrin- β 2-specific CD8⁺ T cells were detected by tetramer staining in tumours and draining lymph nodes (DLNs) from mice challenged with d42m1 parental cells, but not d42m1-es3 cells on day 11 post-transplant. APC, allophycocyanin; PE, phycoerythrin. **b**, Quantification and kinetics of mutant spectrin- β 2 tetramer staining in mice challenged with d42m1 parental cells (*n* = 3, circles) or d42m1-es3 cells (*n* = 3, squares). **c**, Growth of d42m1-es3 tumour cell clones transduced with wild-type (*n* = 5, squares) or mutant spectrin- β 2 (*n* = 5, circles) and control d42m1-es3 cells (*n* = 5, triangles) after transplantation (1×10^6 cells) into wild-type mice. Data are presented as average tumour diameter \pm s.e.m. **d**, d42m1-es3 tumours reconstituted with wild-type (WT.3) or mutant spectrin- β 2 (mu.14) were harvested at day 11 and CD8 α ⁺ T cells were stained with mutant spectrin- β 2 tetramers. **e**, Growth of a mixture of d42m1-T2-RFP (95%) and of d42m1-T3-GFP (5%) after transplantation (1×10^6 total cells) into wild-type (*n* = 5, solid lines, closed squares) or Rag2^{-/-} (*n* = 2, dashed lines, open squares) mice. **f**, Tumour outgrowth in Rag2^{-/-} or wild-type (WT) mice treated or untreated with monoclonal antibodies that deplete CD4⁺ or CD8⁺ T cells after challenge with 1×10^6 cells of a d42m1 mixture (95% d42m1-T2-RFP and 5% d42m1-T3-GFP). Data are presented as per cent tumour positive mice from 2–4 independent experiments (*n* = 2–5 mice per group). **g**, **h**, GFP and RFP expression (**g**) and mutant spectrin- β 2 expression (**h**) were analysed in the d42m1-T2-RFP/d42m1-T3-GFP tumour cell mixture before injection and from tumours that grew out in Rag2^{-/-} mice (RagPass) or escaped in wild-type mice by flow cytometry (**g**) or qRT-PCR (**h**). Data are representative of two independent experiments. Samples were compared using an unpaired, two-tailed Student's *t*-test (**P* < 0.05, ***P* < 0.01, ****P* < 0.001; NS, not significant).

H31m1 reveals that it expresses multiple potential strong neoantigens (19 potential strong binders to H-2D^b and 58 to H-2K^b) (Supplementary Fig. 11a) and induces both H-2D^b- and H-2K^b-restricted

CD8⁺ T-cell responses during rejection (Supplementary Fig. 11b). This result suggests that H31m1 shows an even more complex antigenicity than d42m1 and probably explains why H31m1 never produces escape tumours in wild-type mice (Supplementary Fig. 11c).

Chemically induced tumours have had a critical role in the history of tumour immunology, providing the first unequivocal demonstration of tumour-specific antigens^{26,27} and, subsequently, the first evidence of cancer immunoediting^{1–3}. It is therefore significant that this same model has now provided new insights into the antigenic targets of cancer immunoediting and some of the key molecular mechanisms that drive the process. Although more work is needed to determine whether and how frequently this process occurs during development of spontaneous and carcinogen-induced human cancers, it is tempting to speculate that a genomics approach to tumour antigen identification could, in the future, facilitate the development of individualized cancer immunotherapies directed at tumour-specific—rather than cancer-associated—antigens.

METHODS SUMMARY

d42m1 and H31m1 MCA-induced sarcomas were generated in male 129/Sv Rag2^{−/−} mice as previously described¹. Total RNA was isolated from low-passage MCA-induced sarcoma cell lines and skin fibroblasts from male 129/Sv Rag2^{−/−} mice using the RNeasy Mini kit (Qiagen) and cDNA was prepared using oligo (dT) primers and SuperScript II Reverse Transcriptase (Invitrogen). Illumina libraries prepared with this cDNA were hybridized to biotinylated Agilent mouse exome probes. Library components were captured using streptavidin-coated magnetic beads (DynaBeads), PCR amplified and sequenced using an Illumina GAIIx analyser (cDNA CapSeq). Putative somatic mutations were identified using VarScan 2 (v.2.2.4). Missense mutations were analysed for potential neopeptide binding to MHC class I using an algorithm²⁰ available at Immune Epitope Database and Analysis Resource (<http://www.immuneepitope.org>) and were expressed as affinity values (reciprocal of the predicted IC₅₀ multiplied by 100).

All tumour cell lines were injected subcutaneously in the flank of naïve syngeneic male mice (1 × 10⁶ cells). Ten d42m1 tumour cell clones were isolated from the parental cell line by limiting dilution. Escape tumours of d42m1 were harvested from tumours growing in wild-type mice and cell lines were produced. To generate the C3 d42m1-specific CTL clone, splenocytes from a mouse that rejected d42m1 were harvested, stimulated with parental d42m1 target cells pre-treated with 100 U ml^{−1} IFN-γ for 48 h and irradiated with 100 Gy and cloned by limiting dilution. To clone the antigen recognized by the C3 CTL clone, a d42m1 cDNA library was cloned into pcDNA3 (Invitrogen), transfected into COS cells expressing mouse H-2D^b, and screened for C3 reactivity by IFN-γ ELISA (eBioscience). Mutant spectrin-β2 expression was detected by qRT-PCR using mutation-specific primers. H-2D^b tetramers were generated with 905–913 mutant spectrin-β2 peptides by the NIH Tetramer Facility (Emory).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 17 August; accepted 2 December 2011.

Published online 8 February 2012.

- Shankaran, V. *et al.* IFNγ and lymphocytes prevent primary tumour development and shape tumour immunogenicity. *Nature* **410**, 1107–1111 (2001).
- Dunn, G. P., Bruce, A. T., Ikeda, H., Old, L. J. & Schreiber, R. D. Cancer immunoediting: from immunosurveillance to tumor escape. *Nature Immunol.* **3**, 991–998 (2002).
- Koebel, C. M. *et al.* Adaptive immunity maintains occult cancer in an equilibrium state. *Nature* **450**, 903–907 (2007).
- Vesely, M. D., Kershaw, M. H., Schreiber, R. D. & Smyth, M. J. Natural innate and adaptive immunity to cancer. *Annu. Rev. Immunol.* **29**, 235–271 (2011).
- Schreiber, R. D., Old, L. J. & Smyth, M. J. Cancer immunoediting: integrating immunity's roles in cancer suppression and promotion. *Science* **331**, 1565–1570 (2011).
- Futreal, P. A. *et al.* A census of human cancer genes. *Nature Rev. Cancer* **4**, 177–183 (2004).
- Chen, A. C. & Herschman, H. R. Tumorigenic methylcholanthrene transformants of C3H/10T1/2 cells have a common nucleotide alteration in the c-Ki-ras gene. *Proc. Natl Acad. Sci. USA* **86**, 1608–1611 (1989).

- Tuveson, D. A. *et al.* Endogenous oncogenic K-ras(G12D) stimulates proliferation and widespread neoplastic and developmental defects. *Cancer Cell* **5**, 375–387 (2004).
- Kirsch, D. G. *et al.* A spatially and temporally restricted mouse model of soft tissue sarcoma. *Nature Med.* **13**, 992–997 (2007).
- Ley, T. J. *et al.* DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature* **456**, 66–72 (2008).
- Ding, L. *et al.* Genome remodelling in a basal-like breast cancer metastasis and xenograft. *Nature* **464**, 999–1005 (2010).
- Shah, S. P. *et al.* Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature* **461**, 809–813 (2009).
- Lee, W. *et al.* The mutation spectrum revealed by paired genome sequences from a lung cancer patient. *Nature* **465**, 473–477 (2010).
- Pleasant, E. D. *et al.* A small-cell lung cancer genome with complex signatures of tobacco exposure. *Nature* **463**, 184–190 (2010).
- Totoki, Y. *et al.* High-resolution characterization of a hepatocellular carcinoma genome. *Nature Genet.* **43**, 464–469 (2011).
- Puente, X. S. *et al.* Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. *Nature* **475**, 101–105 (2011).
- Pleasant, E. D. *et al.* A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* **463**, 191–196 (2010).
- Boon, T., Coulie, P. G., Van den Eynde, B. J. & van der Bruggen, P. Human T cell responses against melanoma. *Annu. Rev. Immunol.* **24**, 175–208 (2006).
- Segal, N. H. *et al.* Epitope landscape in breast and colorectal cancer. *Cancer Res.* **68**, 889–892 (2008).
- Nielsen, M. *et al.* Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. *Protein Sci.* **12**, 1007–1017 (2003).
- van der Bruggen, P. *et al.* A gene encoding an antigen recognized by cytolytic T lymphocytes on a human melanoma. *Science* **254**, 1643–1647 (1991).
- Khong, H. T. & Restifo, N. P. Natural selection of tumor variants in the generation of “tumor escape” phenotypes. *Nature Immunol.* **3**, 999–1005 (2002).
- Dunn, G. P., Old, L. J. & Schreiber, R. D. The three Es of cancer immunoediting. *Annu. Rev. Immunol.* **22**, 329–360 (2004).
- DuPage, M., Mazumdar, C., Schmidt, L. M., Cheung, A. F. & Jacks, T. Expression of tumour-specific antigens underlies cancer immunoediting. *Nature* doi:10.1038/nature10803 (this issue).
- Yewdell, J. W. Confronting complexity: real-world immunodominance in antiviral CD8⁺ T cell responses. *Immunity* **25**, 533–543 (2006).
- Prehn, R. T. & Main, J. M. Immunity to methylcholanthrene-induced sarcomas. *J. Natl. Cancer Inst.* **18**, 769–778 (1957).
- Old, L. J. & Boyse, E. A. Immunology of experimental tumors. *Annu. Rev. Med.* **15**, 167–186 (1964).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to J. Archambault for expert technical assistance, T. H. Hansen (Washington University) for providing MHC class I antibodies, S. Horvath and P. M. Allen (Washington University) for synthesizing MHC class I peptides, the National Institutes of Health (NIH) Tetramer Core Facility for producing MHC class I tetramers, and T. S. Stappenbeck (Washington University) for technical help in recovering frozen tumour samples. We also thank E. Unanue, P. M. Allen and J. Bui for criticisms and comments, all members of the Schreiber laboratory for discussions, and the many members of The Genome Institute at Washington University School of Medicine, especially L. Ding for her insights into our analytical approaches. This work was supported by grants to R.D.S. from the National Cancer Institute, the Ludwig Institute for Cancer Research, the Cancer Research Institute, and the WWWW Foundation; and to E.R.M. from the National Human Genome Research Institute. M.D.V. is supported by a pre-doctoral fellowship from the Cancer Research Institute. J.P.A. is supported by the Howard Hughes Medical Institute and the Ludwig Center for Cancer Immunotherapy; M.J.S. by the National Health and Medical Research Council of Australia (NH&MRC) and from the Association for International Cancer Research; and L.J.O. by the Ludwig Institute for Cancer Research and the Cancer Research Institute.

Author Contributions H.M. and M.D.V. were involved in all aspects of this study including planning and performing experiments, analysing and interpreting data, and writing the manuscript. C.G.R., R.U., C.D.A., J.M.W., Y.-S.C. and L.K.S. also performed experiments and analysed data. V.J.M., R.D. and members of The Genome Institute performed Illumina library preparation, cDNA capture and sequencing as well as validation Roche/454 pyrosequencing and 3730 sequencing. D.C.K. analysed and interpreted sequencing data from this study and previously published cancer genome data. J.H. and T.W. analysed cDNA CapSeq data for potential MHC class I epitopes. M.C.W. performed the phylogenetic analysis on the tumour cells. J.P.A., M.J.S. and L.J.O. interpreted data and contributed to the preparation of the final manuscript. E.R.M. and R.D.S. oversaw all the work performed, planned experiments, interpreted data and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to R.D.S. (schreiber@immunology.wustl.edu).

METHODS

Mice. *Ifngr1*^{−/−} mice²⁸ and *Ifnar1*^{−/−} mice²⁹ on a 129/Sv background were originally provided by M. Aguet and were bred in our specific pathogen-free animal facility. Wild-type and *Rag2*^{−/−} mice were purchased from Taconic Farms. All mice were male and on a 129/Sv background and were housed in our specific pathogen-free animal facility. For all experiments, male mice were 8–12 weeks of age and studies were performed in accordance with procedures approved by the AAALAC accredited Animal Studies Committee of Washington University in St. Louis.

Tumour transplantation. MCA-induced sarcomas used in this study were generated in male 129/Sv strain wild-type or *Rag2*^{−/−} mice and banked as low-passage tumour cells as previously described¹. Tumour cells derived from frozen stocks were propagated *in vitro* in RPMI media (Hyclone) supplemented with 10% FCS (Hyclone) and injected subcutaneously in 150 µl of endotoxin-free PBS into the flanks of recipient mice. Tumour cells were >90% viable at the time of injection as assessed by trypan blue exclusion and tumour size was quantified as the average of two perpendicular diameters. For antibody depletion studies, 250 µg of control IgG (PIP), anti-CD4 (GK1.5) or anti-CD8α (YTS169.4) were injected intraperitoneally into mice at day −1 and every 7 days thereafter.

Isolation of normal skin fibroblasts. Skin fibroblasts were isolated from three independent male 129/Sv *Rag2*^{−/−} pups by harvesting skin and incubating in 0.25% trypsin (Hyclone) at 37 °C for 30 min before washing in DMEM media (Hyclone). After washing, chunks of skin were filtered to achieve single-cell suspensions and cultured *in vitro* with DMEM media. After three passages, skin fibroblasts were harvested to isolate genomic DNA and total RNA.

Extraction of genomic or complementary DNA. Genomic DNA from sarcoma cells and normal skin fibroblasts was extracted using DNeasy Blood & Tissue Kit (Qiagen). For cDNA isolation, total RNA from sarcoma cells and normal skin fibroblasts was isolated using RNeasy Mini kit (Qiagen) and cDNA was synthesized using oligo (dT) primers and SuperScript II Reverse Transcriptase (Invitrogen).

cDNA CapSeq. cDNA samples from each tumour (100 ng) were constructed into Illumina libraries according to the manufacturer's protocol (Illumina) with the following modifications. First, cDNA was fragmented using Covaris S2 DNA Sonicator (Covaris) in 1× end-repair buffer followed by the direct addition of the enzyme repair cocktail (Lucigen). Fragment sizes ranged between 100–500 bp. Second, Illumina adaptor-ligated DNA was amplified in four 50 µl PCRs for five cycles using 4 µl adaptor-ligated cDNA, 2× Phusion Master Mix and 250 nM forward and reverse primers, 5'-AATGATACGCGACCGAGATCTACACTCTTCCCTACACGACGCTCTCCGATC and 5'-CAAGCAGAAGACGCATACGAGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC, respectively. Third, Solid Phase Reversible Immobilization (SPRI) bead cleanup was used to purify the PCR-amplified library and to select for 300–500 bp fragments. Five-hundred nanograms of the size-fractionated Illumina library were hybridized with the Agilent mouse exome reagent. After hybridization at 65 °C for 24 h, we added 50 µl of DynaBeads M-270 streptavidin-coated paramagnetic beads (10 mg ml^{−1}) to selectively remove the biotinylated Agilent probes and hybridized cDNA library fragments. The beads were washed according to manufacturer's protocol (Agilent) and the captured library fragments were released into solution using 50 µl of 0.125 N NaOH and neutralized with an equal volume of neutralization buffer (Agilent). The recovered fragments then were PCR amplified according to the manufacturer's protocol using 11 cycles in the PCR. Illumina library quantification was completed using the KAPA SYBR FAST qPCR Kit (KAPA Biosystems). The qPCR result was used to determine the quantity of library necessary to produce 180,000 clusters on a single lane of the Illumina GAIIx. One lane of 100 bp paired-end data was generated for each captured sample (as cDNA was used as the source for sequencing, we refer to this process as cDNA Capture Sequencing or cDNA CapSeq). Illumina reads were aligned to the NCBI build 37 (Mm9) mouse reference sequence using BWA³⁰ v.0.5.5 (with −q 5 soft trimming). Alignments from multiple lanes for the same sample were merged together using SAMtools r599, and duplicates were marked using Picard v.1.29.

Mutation detection and annotation. Putative somatic mutations were identified using VarScan 2 (v.2.2.4)³¹ with the parameters '−min-coverage 3−min-var-freq 0.08−p-value 0.10−somatic-p-value 0.05−strand-filter 1' and specifying a minimum mapping quality of 10. Variants whose supporting reads exhibited read position bias (average read position <10 or >90), strand bias (>99% of reads on one strand), or mapping quality (score difference >30, or mismatch quality sum difference >100) relative to reference supporting reads were removed as probable false positives. We also required that the variant allele be present in at least 10% of tumour reads and no more than 5% of normal reads. The SNVs meeting these criteria were annotated using an internal database of GenBank/Ensembl transcripts (v58_73k). In the event that a variant was annotated using multiple transcripts, the annotation of most severe effect was used. Non-silent coding mutations (missense, nonsense/nonstop or splice site) were prioritized for downstream analysis.

Mutation rate and overlap comparisons. Mutation rates were estimated for each tumour sample using the number of putative 'tier 1' SNVs (missense, nonsense/nonstop, splice site, silent or noncoding RNA). To account for variability in coverage between samples, the SNV count for each tumour sample (S) was divided by a coverage factor (F), computed as the fraction of all tier 1 SNVs identified in any tumour sample ($n = 16,991$) that were covered by at least four reads in a given sample. For example, in the d42m1 parental sample, 15,852 of 16,991 tier 1 SNV positions were covered, for a coverage factor of 93.30%. The number of coverage-adjusted mutations in each sample was divided by the total size of tier 1 space in the mouse genome (43.884 Mbp) to determine the number of coding mutations per megabase (R).

$$R = (S/F) / (43.884 \text{ Mbp})$$

For the mutation overlap comparisons and relatedness-to-parental-tumour analysis, only high-confidence missense mutations were used (that is, 20× or above). A mutation was considered 'shared' between two samples if both samples had a predicted mutation at the same genomic position. For the comparison of mutated genes between d42m1 and H31m1 parental lines, a gene was considered 'shared' if both d42m1 and H31m1 samples had a predicted missense mutation in that gene, even if the mutations did not occur at the same position.

Roche/454 sequencing and validation. PCR primers were designed for 11 SNVs predicted to be somatic in d42m1 tumour samples, as well as 11 control sites that were H31m1-specific, low-confidence, or removed by the false-positive filter. All 22 SNVs were PCR amplified individually in 11 samples (SK1.1, d42m1, H31m1, T2, T3, T5, T9, T10, es1, es2 and es3) using MID-tailed primers to enable sample identification. PCR products were pooled together before sequencing on a quarter run of the Roche/454 Titanium platform. Read sequences and quality scores were extracted from 454 data files using *sffinfo* (454 proprietary software) then aligned to the mouse build 37 reference sequence (Mm19) using SSAHA2 v.2.5.3³² with the SAM output option. Alignments were imported to BAM format and a 'pileup' assembly file generated using SAMtools v.0.1.18³³. The average 454 sequence depth for targeted positions was 1,216× per sample. Validation read counts and allele frequencies in each sample at each variant position were determined using the *pileup2cns* command of VarScan v.2.2.7³¹. At least 20 reads with base quality of 20 or higher were required to confirm or refute a variant. 454 sequencing data and the primers used are presented in Supplementary Table 4.

3730 sequencing and validation. Eight SNVs predicted to be somatic were selected for validation by PCR and 3730 sequencing in flow-sorted CD45⁺ and CD45[−] cells from the original d42m1 tumour. Genomic DNA and cDNA from CD45[−] (tumour) cells, and cDNA from CD45⁺ (normal immune) cells were used for PCR amplification and then PCR products were sequenced individually on ABI 3730 using universal primers. Manual review was performed using amplicon-based assembly in the Integrative Genomics Viewer (IGV)³⁴ to determine the somatic status for each site. Data are presented in Supplementary Table 4.

MHC class I epitope prediction. All missense mutations for each d42m1-related tumour or H31m1 were analysed for the potential to form MHC class I neopeptides that bind to either H-2D^b or H-2K^b molecules. The artificial neural network (ANN) algorithm provided by the Immune Epitope Database and Analysis Resource (<http://www.immuneepitope.org>) was used to predict epitope binding affinities²⁰ and the results were ultimately expressed as affinity values (1/IC₅₀ × 100). Predicted strong affinity epitopes expressed in d42m1 regressor tumours are listed in Supplementary Table 5.

Phylogenetic analysis of tumour samples. Sequencing data from normal *Rag2*^{−/−} fibroblasts, d42m1 parental cells, d42m1 regressor clones, d42m1 regressor clones, d42m1 escape tumours and H31m1 tumour cells were compared using PHYLogeny Inference Package (PHYLP)³⁵ to generate a phylogenetic tree displaying the relatedness of each sample.

Antibodies. Anti-H-2K^b (B8-24-3) and anti-H-2D^b (B22/249) monoclonal antibodies were provided by T. H. Hansen (Washington University School of Medicine). Anti-CD4 (GK1.5), anti-CD8α (YTS169.4) monoclonal antibodies and control immunoglobulin (PIP, a monoclonal antibody specific for bacterial glutathione S-transferase) were produced from hybridoma supernatants and purified in endotoxin-free form by Protein G affinity chromatography (Leinco Technologies). Purified Rat IgG was purchased from Sigma (St. Louis). CD45-FITC, CD45-PE, CD8-APC and purified anti-CD16/32 were purchased from BioLegend.

cDNA library construction and screening. To generate a d42m1 tumour cell cDNA library, mRNA was isolated from parental d42m1 tumour cells using a QuickPrep mRNA Purification kit (Amersham), converted into cDNA using SuperScript II First Strand Synthesis System (Invitrogen) and inserted into the EcoRI site of the expression vector pcDNA3 (Invitrogen). The cDNA library was divided into pools of 100 bacterial colonies with 200–300 ng of DNA from each

pool transfected into 2.5×10^4 monkey COS cells engineered to ectopically express mouse H-2D^b (COS-D^b) cells using Lipofectamine 2000. After 48 h, 5×10^3 C3 CTL cells were added, and supernatants were assayed for IFN- γ release 24 h later by ELISA. A single positive cDNA clone was isolated after screening 120,000 cDNA colonies. The putative H-2D^b-binding peptide VAVVNQIAL was predicted using the algorithm available at the Immune Epitope Database and Analysis Resource, <http://www.immuneepitope.org/>. The peptides were produced by P. Allen and S. Horvath (Washington University School of Medicine).

Expression vectors. Full-length cDNA encoding wild-type spectrin- β 2 and mutant spectrin- β 2 were cloned from parental d42m1 tumour cells by RT-PCR using primer pairs 5'-TGAGACAGTCAAGATGACGACCAGGTAGCCACA-3' and 5'-CGGGACAACAGGGAAGTTCCTTCTTCTGCCGA-3'. Wild-type and mutant spectrin- β 2 cDNA were subcloned from the TOPO-XL vector (Invitrogen) into the retrovirus (RV)-GFP vector³⁶. To generate the RV-RFP vector, full-length cDNA encoding RFP was cloned from the pTurboRFP-C vector (Evrogen) by RT-PCR using primer pairs 5'-ATCTCAGAATTCATGAGC GAGCTGATCAAGGA-3' and 5'-ATCTCAGGATCCTTATCTGTGCCCA GTTTGCTAG-3'. RFP cDNA was then cloned into the RV vector. To remove candidate T-cell epitopes in RFP, the nucleotide A was replaced by G at position 334 in the cDNA, resulting in amino acid substitution N112D. Coding sequences of the constructs were verified by DNA sequencing (Big Dye method; Applied Biosciences). The dominant-negative version of the IFNGR1 subunit (IFNGR1 Δ IC) was expressed into H31m1 and d42m1 tumour cells as previously described³⁷.

Establishment of CTL lines and clones. To generate the d42m1-specific C3 CTL clone, wild-type mice were injected with 1×10^6 parental d42m1 tumour cells. Fourteen days later, the spleen was harvested from a mouse that rejected the tumour and a CTL line was established by stimulating 40×10^6 splenocytes with 2×10^6 parental d42m1 tumour cells pre-treated for 48 h with 100 U ml^{-1} of recombinant murine IFN- γ and irradiated (100 Gy). After CD8⁺ T-cell purification using magnetic beads (Miltenyi Biotec) and limiting dilution, the CTL clone C3 was obtained.

Measurement of IFN- γ production. To generate target cells, tumour cells were treated with 100 U ml^{-1} IFN- γ for 48 h and irradiated with 100 Gy before use. The C3 CTL clone was co-cultured at the indicated ratios with target tumour cells (10,000 or 5,000 cells) in 96-well round-bottomed plates overnight. IFN- γ in supernatants was quantified using an IFN- γ ELISA kit (eBioscience). For blocking assays, $10 \mu\text{g ml}^{-1}$ of anti-CD8 (YTS-169.4), anti-CD4 (GK1.5) or control immunoglobulin (PIP) were added to the cell culture of effector (C3 CTL clone) and target cells (tumours).

Cytotoxicity assay. To generate target cells, tumour cells were treated with 100 U ml^{-1} rMuIFN- γ for 48 h before use. One million tumour cells were labelled with $25 \mu\text{Ci}$ of $\text{Na}_2^{51}\text{CrO}_4$ (PerkinElmer) for 90 min at 37 °C, washed and 10,000 cells seeded per well in 96-well round-bottom plates. The C3 CTL clone was co-cultured with the tumour target cells at the indicated effector/target cell ratios and incubated for 4 h at 37 °C in 5% CO_2 . Radioactivity was detected in the supernatants and per cent specific killing was defined as (experimental condition c.p.m. – spontaneous c.p.m.) / (maximal (detergent) c.p.m. – spontaneous c.p.m.) $\times 100$. Data points were obtained in duplicate.

Fluorescence-activated cell sorting analysis. For flow cytometry, cells were stained for 20 min at 4 °C with 500 ng of Fc block (anti-CD16/32) and 200 ng of CD45, CD4 or CD8 α in 100 μl of staining buffer (PBS with 1% FCS and 0.05% NaN_3 (Sigma)). Propidium iodide (PI) (Sigma) was added at $1 \mu\text{g ml}^{-1}$ immediately before FACS analysis. For quantitative analysis of tumour-infiltrating lymphocytes/leukocytes (TIL) and lymph node populations, a CD45⁺ PI⁺ gate was used and gated events were collected on a FACSCalibur (BD Biosciences) and analysed using FloJo software.

Tumour, draining lymph node and spleen harvest. After tumour cell transplantation, established tumours were excised from mice, minced and treated with 1 mg ml^{-1} type IA collagenase (Sigma) in HBSS (Hyclone) for 2 h at room temperature (22 °C). The ipsilateral inguinal tumour draining lymph nodes and spleen were also harvested and crushed between two glass slides and vigorously resuspended to make single-cell suspensions.

Tetramers. H-2D^b tetramers conjugated to PE were prepared with mutant spectrin- β 2 peptides and produced by the NIH Tetramer Core Facility (Emory University).

Mutation-specific RT-PCR and real-time RT-PCR. Total RNA from tumour cells was isolated by RNeasy Mini kit (Qiagen) and cDNA was synthesized from the total RNA using oligo (dT) primers and SuperScript II Reverse Transcriptase (Invitrogen). Real-time PCR specific for wild-type spectrin- β 2, mutant spectrin- β 2 and GAPDH using the SYBR Green Mastermix kit (Applied Biosystems) were performed on ABI 7000. The primer sequences for used for mutant spectrin- β 2 are 5'-GGTGAACCAGATTGCACT-3' and 5'-TGTCCACCAGTTCTCTGAAC-3'.

Detection of mutation in spectrin- β 2 cDNA. The point mutation in the spectrin- β 2 gene creates a PstI restriction site (CGGCAG to CTGCAG, underlined letters indicate the site of mutation). To amplify spectrin- β 2 cDNA we used a forward primer (ACCCTGGCCCTGTACAAGAT) and reverse primer (TAGACTCGATGACCTTGGTCT). The PCR conditions used were 94 °C for 2 min, followed by 35 cycles of 94 °C for 30 s, 55 °C for 30 s and 72 °C for 30 s. The PCR products were digested for 2 h at 37 °C with PstI restriction enzyme, which cleaved mutant spectrin- β 2, but not wild-type spectrin- β 2, and generates a 200 bp fragment from cDNA. The products were resolved by electrophoresis on a 1.2% agarose gel and visualized by ethidium bromide staining.

Isolation of non-transformed cells from d42m1 biopsy. A frozen d42m1 tumour biopsy from the original d42m1 tumour was thawed and treated with 1 mg ml^{-1} type IA collagenase (Sigma) in HBSS for 2 h at room temperature. After filtration, single-cell suspensions were stained for 20 min at 4 °C with 500 ng of Fc block (anti-CD16/32) and 200 ng of CD45-PE in 100 μl of staining buffer. Propidium iodide was added at $1 \mu\text{g ml}^{-1}$ immediately before sorting. A CD45⁺ PI⁺ gate was used and the top 15% and the bottom 15% of gated events were collected using a FACSAria II (BD Biosciences). Sorted CD45⁺ cells (host leukocytes) and CD45⁺ cells (primary d42m1 tumour cells) were collected and genomic DNA as well as RNA was isolated to synthesize cDNA for 3730 sequencing to validate that the mutation calls detected by Illumina were somatic and tumour specific.

Statistical analysis. Samples were compared using an unpaired, two-tailed Student's *t*-test, unless specified.

28. Huang, S. *et al.* Immune response in mice that lack the interferon- γ receptor. *Science* **259**, 1742–1745 (1993).
29. Muller, U. *et al.* Functional role of type I and type II interferons in antiviral defense. *Science* **264**, 1918–1921 (1994).
30. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
31. Koboldt, D. C. *et al.* VarScan: variant detection in massively parallel sequencing of individual and pooled samples. *Bioinformatics* **25**, 2283–2285 (2009).
32. Ning, Z., Cox, A. J. & Mullikin, J. C. SSAHA: a fast search method for large DNA databases. *Genome Res.* **11**, 1725–1729 (2001).
33. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
34. Robinson, J. T. *et al.* Integrative genomics viewer. *Nature Biotechnol.* **29**, 24–26 (2011).
35. Felsenstein, J. Phylogeny Inference Package. *Cladistics* **5**, 164–166 (1989).
36. Ranganath, S. *et al.* GATA-3-dependent enhancer activity in *IL-4* gene regulation. *J. Immunol.* **161**, 3822–3826 (1998).
37. Dighe, A. S., Richards, E., Old, L. J. & Schreiber, R. D. Enhanced *in vivo* growth and resistance to rejection of tumor cells expressing dominant negative IFN γ receptors. *Immunity* **1**, 447–456 (1994).

Expression of tumour-specific antigens underlies cancer immunoediting

Michel DuPage¹, Claire Mazumdar¹, Leah M. Schmidt¹, Ann F. Cheung¹ & Tyler Jacks^{1,2}

Cancer immunoediting is a process by which immune cells, particularly lymphocytes of the adaptive immune system, protect the host from the development of cancer and alter tumour progression by driving the outgrowth of tumour cells with decreased sensitivity to immune attack^{1,2}. Carcinogen-induced mouse models of cancer have shown that primary tumour susceptibility is thereby enhanced in immune-compromised mice, whereas the capacity for such tumours to grow after transplantation into wild-type mice is reduced^{2,3}. However, many questions about the process of cancer immunoediting remain unanswered, in part because of the known antigenic complexity and heterogeneity of carcinogen-induced tumours⁴. Here we adapted a genetically engineered, autochthonous mouse model of sarcomagenesis to investigate the process of cancer immunoediting. This system allows us to monitor the onset and growth of immunogenic and non-immunogenic tumours induced *in situ* that harbour identical genetic and histopathological characteristics. By comparing the development of such tumours in immune-competent mice with their development in mice with broad immunodeficiency or specific antigenic tolerance, we show that recognition of tumour-specific antigens by lymphocytes is critical for immunoediting against sarcomas. Furthermore, primary sarcomas were edited to become less immunogenic through the selective outgrowth of cells that were able to escape T lymphocyte attack. Loss of tumour antigen expression or presentation on major histocompatibility complex I was necessary and sufficient for this immunoediting process to occur. These results highlight the importance of tumour-specific-antigen expression in immune surveillance, and potentially, immunotherapy.

To determine whether T lymphocytes influence tumour development, we adapted a mouse model of human soft tissue sarcomagenesis driven by Cre/LoxP-regulated expression of oncogenic K-ras^{G12D} and deletion of p53 to allow for the control of tumour immunogenicity⁵. Sarcomas were induced in either immune-competent *Kras*^{LSL-G12D/+}; *p53*^{fl/fl}; *Rag2*^{+/-} (KP) or lymphocyte-deficient *Kras*^{LSL-G12D/+}; *p53*^{fl/fl}; *Rag2*^{-/-} (KPR) mice by intramuscular injection of lentiviral vectors that expressed Cre recombinase alone (Lenti-x). To induce sarcomas with potentially immunogenic antigens, we used vectors that also expressed the T-cell antigens SIYRYYL (SIY) and two antigens from ovalbumin (SIINFEKL (SIN), OVA₂₅₇₋₂₆₄) and OVA₃₂₃₋₃₃₉) fused to the carboxy terminus of luciferase (Lenti-LucOS). Intramuscular injection of Lenti-LucOS led to tumour formation in 100% of KPR mice but only 27% of KP mice by 140 days (Fig. 1a, $P < 0.0001$). Additional sarcomas ultimately developed in KP mice but with dramatically delayed kinetics (latency of 194.8 ± 43.4 days) compared with KPR mice (73.6 ± 4.3 days) (Fig. 1c, $P < 0.02$). We also observed a difference in the penetrance of sarcoma development in KPR versus KP mice by 140 days with Lenti-x (89% versus 43%, respectively), although the difference was less dramatic than observed with Lenti-LucOS (Fig. 1b, $P < 0.0005$). This suggests that in this model, tumour immunosurveillance may not necessitate the introduction of highly immunogenic tumour-specific

antigens (TSAs). The observed immunosurveillance against Lenti-x tumours could result from the lentiviral infection required to induce tumours, the acquisition of TSAs during tumour development, or the immunogenicity of Cre itself. However, in a previous study, we found that Cre was not highly immunogenic when expressed in developing lung adenocarcinomas⁶. Although Lenti-x-induced sarcoma development was slightly delayed in immune-competent (KP) mice (114.9 days in KP versus 79.5 days in KPR mice), it was not significant (Fig. 1c, $P = 0.11$). The increased latency that is specific to Lenti-LucOS tumours may be the result of an equilibrium between replicating tumour cells and T cells that recognize antigens expressed from the LucOS vector and restrain tumour progression^{1,7}.

Rag2 (recombination activating gene 2) deficiency prevents both T and B lymphocyte development and, therefore, could have pleiotropic effects on the immune response to tumour antigens. To specifically test

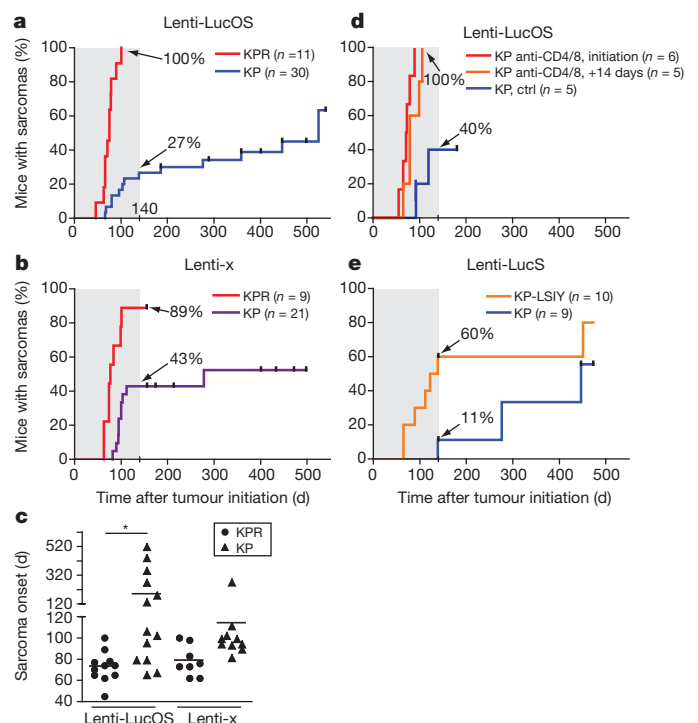


Figure 1 | Sarcoma formation in immunodeficient compared to wild-type mice occurs with increased penetrance and reduced latency. a, b, KPR or KP mice were injected intramuscularly with Lenti-LucOS (a) or Lenti-x (b) and the onset of palpable sarcomas was monitored. c, Time for palpable tumour formation with Lenti-LucOS or Lenti-x in KPR (circles) or KP (triangles) mice. d, Sarcoma formation in KP mice either untreated or treated with anti-CD4 and anti-CD8 antibodies beginning coincident with or 14 days after Lenti-LucOS injection. e, Sarcoma onset after injection of KP-LSIY or KP littermates with Lenti-LucOS. The percentage of total mice (n) with sarcomas by 140 days (grey boxes) is indicated.

¹Koch Institute for Integrative Cancer Research and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. ²Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

the significance of T-cell responses, we treated mice with antibodies against CD4 and CD8 to deplete T cells concurrent with, or subsequent to, intramuscular injection of Lenti-LucOS. T-cell depletion at tumour initiation, or even 14 days after tumour initiation, led to sarcoma development with complete penetrance and early onset similar to KPR mice (Fig. 1d, $P = 0.001$ and $P = 0.013$ compared to untreated, respectively). To specifically test the importance of $CD8^+$ T cells that recognize the model TSAs, we made use of a regulatable luciferase–SIY fusion gene engineered into the murine Rosa26 locus ($R26^{LSL-SIY}$)⁸. These mice develop specific tolerance to luciferase and SIY due to weak thymic expression and deletion of reactive T cells (Supplementary Fig. 1)⁸. $Kras^{LSL-G12D/+}; p53^{fl/fl}; R26^{LSL-SIY/+}$ (KP-LSIY) mice injected with Lenti-LucS, a lenti-vector that expresses Cre and SIY fused to luciferase, were more susceptible to sarcoma formation and developed tumours earlier than KP littermates (Fig. 1e, $P = 0.058$). Thus, lymphocyte-mediated protection from sarcoma formation requires $CD8^+$ T cells that respond to non-self antigens expressed in tumours.

A key advantage of this conditional, genetically engineered cancer model over carcinogen-induced models is the capacity to track endogenous T cells specific for tumour antigens during primary tumour development. We used SIY and SIN loaded MHCI/K^b reagents to track tumour-reactive $CD8^+$ T cells by flow cytometry. Only mice with Lenti-LucOS sarcomas harboured $CD8^+$ T cells specific to SIY and SIN in the lymph nodes nearest the tumour site as well as in the spleen (Fig. 2a, b). These $CD8^+$ T cells appeared to be completely functional because they produced both IFN- γ and TNF- α upon stimulation

(Fig. 2a–d). Interestingly, this contrasts sharply with results from an analogous model of lung adenocarcinoma in which the activity of T cells responding to the same tumour antigens was very weak, suggesting that different tumour types may use different mechanisms to escape immune attack⁶. We also investigated whether KP mice that did not develop sarcomas after injection with Lenti-LucOS harboured antigen-specific T cells, because such T cells could have protected these mice from sarcoma development. Indeed, we detected fully functional antigen-specific T cells in these mice (Fig. 2c, d and Supplementary Fig. 1), demonstrating that T cells specific to these model TSAs are functional and probably provide significant protection against the development of Lenti-LucOS sarcomas.

Pivotal experiments using methylcholanthrene (MCA)-induced sarcomas revealed that tumours generated in immune-compromised mice, and thus not immunodeficient, are more susceptible to rejection upon transplantation into immune-competent mice². To assay whether autochthonous sarcomas driven by targeted genetic mutations would also display an unedited phenotype, we transplanted independently derived sarcomas from KPR or KP mice into either wild-type or $Rag2^{-/-}$ mice. Whereas freshly isolated Lenti-LucOS-induced tumours (or cell lines) generated in KP mice grew similarly upon transplantation into either wild-type or $Rag2^{-/-}$ mice, most Lenti-LucOS tumours generated in KPR mice were rejected (1/7) or had significantly delayed growth (4/7) (Fig. 3a, b and Supplementary Fig. 2). These results recapitulate the original findings from carcinogen-induced sarcomas in a genetically engineered mouse model of sarcomagenesis.

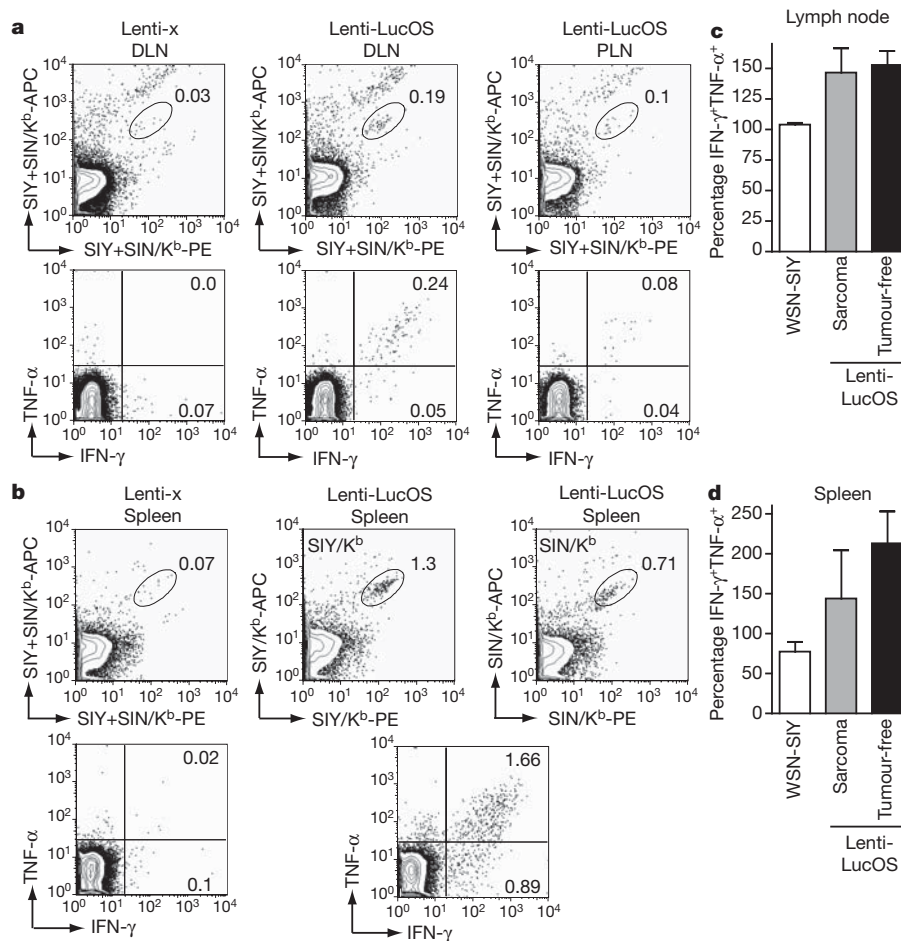


Figure 2 | Functional T-cell responses are generated against antigens expressed in sarcomas. **a**, Top: percentage of gated $CD8^+$ cells specific for SIY and SIN in the inguinal lymph nodes either draining (DLN) or peripheral to (PLN) Lenti-x or Lenti-LucOS tumours. Bottom: IFN- γ and TNF- α cytokine production in SIY+SIY-stimulated $CD8^+$ T cells from mice analysed above. **b**, Analysis of splenocytes as in **a**. **c**, **d**, Cumulative data depicting the percentage

of SIY- and SIN-specific T cells that were IFN- γ ⁺TNF- α ⁺ from lymph nodes (**c**) or spleens (**d**) of KP mice infected with Lenti-LucOS that developed a 'sarcoma' or were 'tumour-free' at 170 days. T cells reactive to SIY were analysed four months after challenge with WSN-SIY (influenza strain expressing SIY). Data represent analysis of 3–4 mice per group, mean \pm s.e.m.

Next we wanted to determine whether Lenti-x-induced sarcomas, which lack the strong T-cell antigens from LucOS, would yield similar results. Interestingly, Lenti-x tumours generated in KPR or KP mice grew equally well when transplanted into wild-type or *Rag2*^{-/-} mice (Fig. 3c, d). It is noteworthy that while autochthonous tumours initiated by Lenti-x appeared partially inhibited by an adaptive immune response (Fig. 1b), in the context of transplantation, we found no evidence of immunoediting (Fig. 3c). This difference may be due to Rag-dependent innate immune cells (NKT and $\gamma\delta$ T cells) that recognize stress or inflammatory ligands. These cells may be sufficient to eliminate a limited number of nascent tumour cells in the context of transformation by lentiviral infection, but not in response to the transplantation of

fully developed tumours^{1,9}. Nevertheless, we suggest that Lenti-x sarcomas from KPR mice grew unabated after transplantation into KP mice because immunoediting by T lymphocytes requires potent TSAs, which these tumours lack. The observed immunogenicity of carcinogen-induced sarcomas generated in immune-compromised mice may be due to the *de novo* generation of potent tumour neoantigens during transformation with mutagens^{9–12}. Importantly, in another study reported in this issue¹³, somatically mutated spectrin- β 2 in a MCA-induced sarcoma was found to act as a potent neoantigen that drove the immunoediting process. In an attempt to introduce immunogenic mutations in Lenti-x tumours, we treated cell lines from these tumours with MCA *in vitro*. Interestingly, such treatment rarely yielded clones with increased immunogenicity (Supplementary Fig. 3). This may indicate that although carcinogens can produce mutations that are immunogenic, it may be a rare event.

If cancer immunoediting by lymphocytes requires potent TSAs, then Lenti-LucOS-induced tumours that appear edited after forming in KP mice may have evaded an immune response by the selective outgrowth of cells lacking these potent antigens^{14–16}. To assess antigen expression, we measured luciferase activity in tumours. Whereas tumours from KPR mice were universally luciferase positive, tumours from KP mice had drastically reduced luciferase activity in all but one of six sarcomas (Fig. 4a, b). Interestingly, this sarcoma had significantly reduced expression of H-2K^b, the MHC class I allele responsible for presenting the SIY and SIN antigens (Fig. 4c). Sarcomas from KP mice treated with anti-CD4 and anti-CD8 antibodies at tumour initiation also retained luciferase activity (5/6 sarcomas luc⁺, Fig. 4a). However, fewer sarcomas retained luciferase expression when mice were treated with anti-CD4 and anti-CD8 antibodies beginning 14 days after tumour initiation (1/5 sarcomas luc⁺), suggesting that immunoediting can occur very early during sarcoma development. Thus, by selectively eliminating cells that express potent TSAs, T lymphocytes drive the escape of tumour cells that either do not express potent antigens or cannot present the antigens to reactive T cells.

In a similar fashion to the antigen loss observed in autochthonous sarcomas, Lenti-LucOS-induced sarcomas from KPR mice lost antigen expression when transplanted into wild-type mice (Supplementary Fig. 4). Importantly, tumours that lost antigen expression after being passed through wild-type mice grew comparably upon secondary transplantation into wild-type and *Rag2*^{-/-} mice, whereas tumours passed through *Rag2*^{-/-} mice did not (Supplementary Fig. 4). To test whether antigen loss was sufficient to provide a means of escape for Lenti-LucOS sarcomas generated in KP mice, we reintroduced the LucOS antigens into sarcomas that had lost expression of the antigens after passage through wild-type mice (referred to as antigen⁻ tumours). Indeed, re-expression of LucOS led to severely reduced tumour growth (Fig. 4d), indicating that loss of antigen expression was the primary means of tumour escape in this setting.

Epigenetic silencing of tumour antigen expression via DNA methylation could be responsible for antigen loss and tumour escape^{17,18}. To test this hypothesis, we treated cell lines that had lost luciferase expression after transplantation into immune-competent mice with 5-aza-2'-deoxycytidine (Aza), which reverses epigenetic silencing by inhibiting DNA methylation. In several lines tested, luciferase activity was restored with Aza treatment (Fig. 4e). Therefore, epigenetic silencing of tumour antigens may represent an important mechanism by which tumours can be edited in response to immune surveillance.

Here we have overcome many of the obstacles of carcinogen-induced models of cancer by using an autochthonous, genetically engineered model of sarcomagenesis to show that T lymphocyte-driven tumour antigen loss is a critical means by which cancer immunoediting occurs in a primary tumour setting. Although this study was limited to investigating the role of anti-tumour immunity by T cells, this model could be adapted to investigate the role of other critical immune cells in cancer immunoediting, such as B cells or NK cells, by either introducing surface-expressed or stress-related antigens into

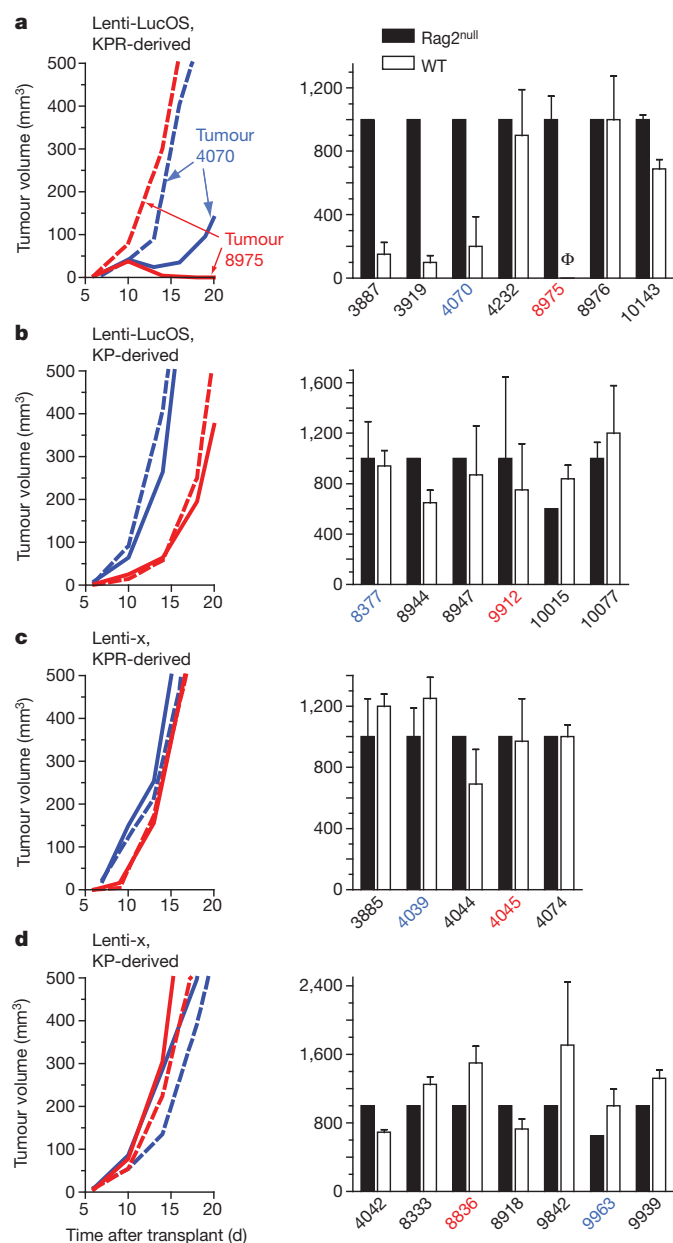


Figure 3 | Cancer immunoediting phenotypes require the presence of potent T-cell antigens. Transplanted tumour growth of Lenti-LucOS-induced sarcomas generated in KPR (a) or KP (b) mice and Lenti-x-induced sarcomas generated in KPR (c) or KP (d) mice. Left column, representative tumour growth curves from two different primary tumours (coloured red or blue) after transplantation into *Rag2*^{null} (dashed lines) or wild-type (WT, solid lines) mice. Right column, comparison of the mean tumour volume \pm s.e.m. for all tumours transplanted. Φ indicates no detectable mass. See Supplementary Fig. 2 for growth curves of tumour lines.

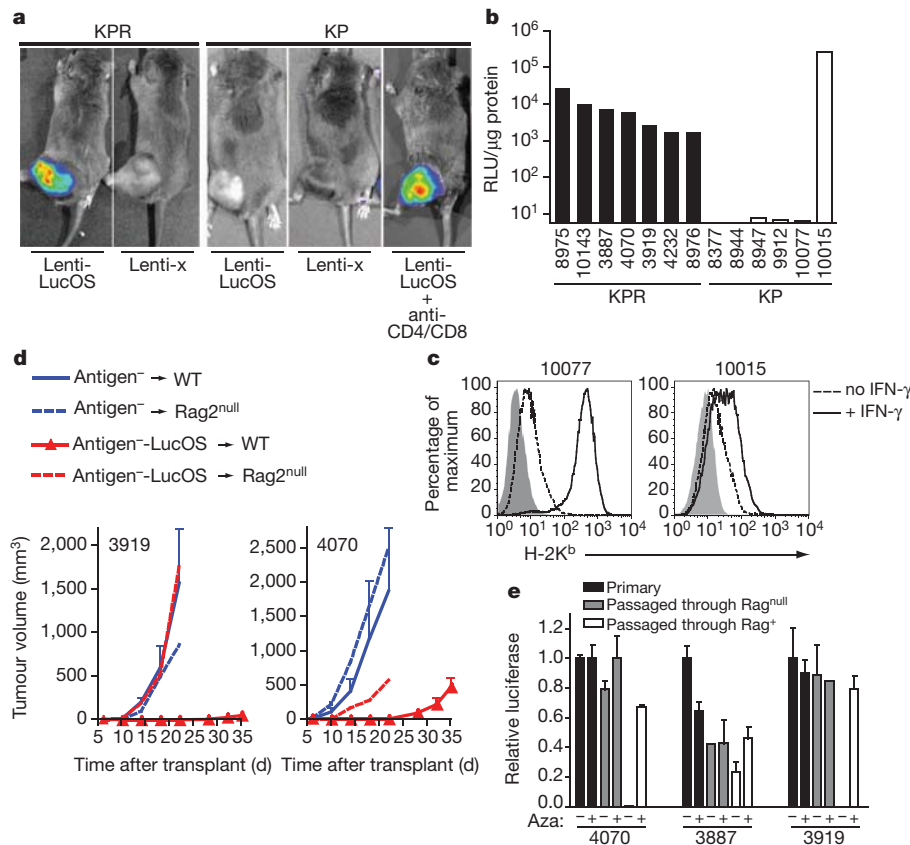


Figure 4 | Immunoeediting occurs by selecting for tumour cells that do not express targeted antigens. **a**, Representative luciferase activity of Lenti-LucOS and Lenti-x-induced sarcomas in KPR, KP or anti-CD4/CD8 treated KP mice. **b**, Luciferase expression in Lenti-LucOS-induced sarcoma cell lines generated in KPR or KP mice. **c**, Freshly harvested sarcomas cultured with IFN- γ (solid line) or untreated (dashed line) were analysed for H-2K^b surface expression (shaded, control antibody). **d**, Growth of two independent tumours (3919 and 4070) that had lost antigen expression (antigen⁻, blue lines) or the same

tumour lines after reintroduction of LucOS (antigen⁻-LucOS, red lines). Mean tumour volume \pm s.e.m. after transplantation into three wild-type mice (solid lines) or one Rag2^{null} mouse (dashed lines). **e**, Relative luciferase activity (compared to the primary sarcoma) \pm 5-aza-2'-deoxycytidine (Aza) of Lenti-LucOS sarcomas from KPR mice (Primary, black columns) that were passaged through Rag2^{-/-} (Passed through Rag^{null}, grey columns) or wild-type mice (Passed through Rag⁺, white columns). Mean \pm s.e.m. from two experiments.

tumours, respectively^{19–21}. This study resulted in two key discoveries. First, oncogene-driven, endogenous tumours can undergo immunoediting in a manner similar to carcinogen-driven tumours if engineered to express model TSAs. The immunogenicity of MCA-induced sarcomas is well-documented, and may be a direct consequence of TSAs that arise from carcinogen-induced mutations of normal genes during tumour development^{9,11–13}. In contrast, cancers that arise spontaneously or by targeted genetic mutations in mice have been reported to be weakly immunogenic^{22–25}. However, the mutational requirements for tumorigenesis in humans may be greater than in mice²⁶, and thus it is possible that spontaneous or genetically engineered mouse models of cancer might underestimate the mutational and antigenic load of most human cancers. This idea is supported by the second critical finding of this study—that tumour immunogenicity is not a universal characteristic of cancer development. By obviating the need for carcinogens, we could induce sarcomas that potentially lacked potent TSAs. These tumours had significantly reduced immunogenicity despite no previous engagement with the adaptive immune system and hence no opportunity for immunoediting. These results provide the first (to our knowledge) experimental system to unify the apparently conflicting results obtained using either carcinogen-induced or genetically targeted mouse models of cancer by identifying TSAs as the critical determinants that invoke adaptive immunosurveillance and immunoediting^{2,22}. We propose that identifying and characterizing TSAs in human cancers may be critical for the generation of more effective anti-cancer immunotherapies in patients suffering from this disease.

METHODS SUMMARY

Experiments used mice of the 129S₄/SvJae strain. All animal studies and procedures were approved by the Massachusetts Institute of Technology's Committee for Animal Care. Sarcomas were induced in KP and KPR mice by intramuscular injection of the hind limb with replication-incompetent lentiviruses expressing Cre recombinase as reported previously^{5,6}. To deplete T cells, anti-CD4 (GK1.5) and anti-CD8 (YTS169.4) antibodies were administered at a dose of 250 μ g per mouse by i.p. injection once weekly for the duration of the experiment. Flow cytometry was performed as described⁶. For transplantation experiments, 2×10^5 freshly isolated tumour cells or cultured tumour cells were transplanted subcutaneously into immune-competent or Rag2^{-/-} mice of the 129S₄/SvJae background. Tumour volumes were calculated by multiplying the length \times width \times height of each tumour. To detect luciferase activity, freshly explanted tumours or cell lines were lysed, mixed with Luciferin reagent (Promega), and relative light units (RLU) were detected with a luminometer (MGM Instruments). Aza treatment used 1 μ M 5-aza-2'-deoxycytidine for three days. *In vivo* bioluminescence images were acquired with the NightOWLII LB983 (Berthold Technologies) or the IVIS Spectrum (Xenogen) after intraperitoneal injection of 1.5 mg beetle luciferin (Promega). Statistical analyses used unpaired two-tailed Fisher exact probability tests or Student's *t*-tests.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 17 August; accepted 16 December 2011.

Published online 8 February 2012.

- Dunn, G. P., Bruce, A. T., Ikeda, H., Old, L. J. & Schreiber, R. D. Cancer immunoediting: from immunosurveillance to tumor escape. *Nature Immunol.* **3**, 991–998 (2002).

2. Shankaran, V. *et al.* IFN γ and lymphocytes prevent primary tumour development and shape tumour immunogenicity. *Nature* **410**, 1107–1111 (2001).
3. Swann, J. B. *et al.* Demonstration of inflammation-induced cancer and cancer immunoediting during primary tumorigenesis. *Proc. Natl Acad. Sci. USA* **105**, 652–656 (2008).
4. Qin, Z. & Blankenstein, T. A cancer immunosurveillance controversy. *Nature Immunol.* **5**, 3–4 (2004); author reply **5**, 4–5 (2004).
5. Kirsch, D. G. *et al.* A spatially and temporally restricted mouse model of soft tissue sarcoma. *Nature Med.* **13**, 992–997 (2007).
6. DuPage, M. *et al.* Endogenous T cell responses to antigens expressed in lung adenocarcinomas delay malignant tumor progression. *Cancer Cell* **19**, 72–85 (2011).
7. Koebel, C. M. *et al.* Adaptive immunity maintains occult cancer in an equilibrium state. *Nature* **450**, 903–907 (2007).
8. Cheung, A. F., Dupage, M. J., Dong, H. K., Chen, J. & Jacks, T. Regulated expression of a tumor-associated antigen reveals multiple levels of T-cell tolerance in a mouse model of lung cancer. *Cancer Res.* **68**, 9459–9468 (2008).
9. Khong, H. T. & Restifo, N. P. Natural selection of tumor variants in the generation of “tumor escape” phenotypes. *Nature Immunol.* **3**, 999–1005 (2002).
10. Prehn, R. T. & Main, J. M. Immunity to methylcholanthrene-induced sarcomas. *J. Natl. Cancer Inst.* **18**, 769–778 (1957).
11. Dubey, P. *et al.* The immunodominant antigen of an ultraviolet-induced regressor tumor is generated by a somatic point mutation in the DEAD box helicase p68. *J. Exp. Med.* **185**, 695–705 (1997).
12. Monach, P. A., Meredith, S. C., Siegel, C. T. & Schreiber, H. A unique tumor antigen produced by a single amino acid substitution. *Immunity* **2**, 45–59 (1995).
13. Matsushita, H. *et al.* Cancer exome analysis reveals a T-cell-dependent mechanism of cancer immunoediting. *Nature* <http://dx.doi.org/10.1038/nature10755> (this issue).
14. Uyttenhove, C., Maryanski, J. & Boon, T. Escape of mouse mastocytoma P815 after nearly complete rejection is due to antigen-loss variants rather than immunosuppression. *J. Exp. Med.* **157**, 1040–1052 (1983).
15. Zhou, G., Lu, Z., McCadden, J. D., Levitsky, H. I. & Marson, A. L. Reciprocal changes in tumor antigenicity and antigen-specific T cell function during tumor progression. *J. Exp. Med.* **200**, 1581–1592 (2004).
16. Stauss, H. J., Van Waes, C., Fink, M. A., Starr, B. & Schreiber, H. Identification of a unique tumor antigen as rejection antigen by molecular cloning and gene transfer. *J. Exp. Med.* **164**, 1516–1530 (1986).
17. Guo, Z. S. *et al.* De novo induction of a cancer/testis antigen by 5-aza-2'-deoxycytidine augments adoptive immunotherapy in a murine tumor model. *Cancer Res.* **66**, 1105–1113 (2006).
18. Güre, A. O., Wei, I. J., Old, L. J. & Chen, Y. T. The SSX gene family: characterization of 9 complete genes. *Int. J. Cancer* **101**, 448–453 (2002).
19. Guerra, N. *et al.* NKG2D-deficient mice are defective in tumor surveillance in models of spontaneous malignancy. *Immunity* **28**, 571–580 (2008); correction **28**, 723 (2008).
20. Schietinger, A. *et al.* A mutant chaperone converts a wild-type protein into a tumor-specific antigen. *Science* **314**, 304–308 (2006).
21. Zitvogel, L., Tesniere, A. & Kroemer, G. Cancer despite immunosurveillance: immunoselection and immunosubversion. *Nature Rev. Immunol.* **6**, 715–727 (2006).
22. Willmsky, G. & Blankenstein, T. Sporadic immunogenic tumours avoid destruction by inducing T-cell tolerance. *Nature* **437**, 141–146 (2005).
23. Embleton, M. J. & Heidelberger, C. Antigenicity of clones of mouse prostate cells transformed *in vitro*. *Int. J. Cancer* **9**, 8–18 (1972).
24. Hewitt, H. B., Blake, E. R. & Walder, A. S. A critique of the evidence for active host defence against cancer, based on personal studies of 27 murine tumours of spontaneous origin. *Br. J. Cancer* **33**, 241–259 (1976).
25. Scott, O. C. Tumor transplantation and tumor immunity: a personal view. *Cancer Res.* **51**, 757–763 (1991).
26. Rangarajan, A. & Weinberg, R. A. Opinion: Comparative biology of mouse versus human cells: modelling human cancer in mice. *Nature Rev. Cancer* **3**, 952–959 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. M. Winslow and A. G. DuPage for critical reading of this manuscript. This work was supported by grant 1 U54 CA126515-01 from the NIH, and partially by Cancer Center Support (core) grant P30-CA14051 from the National Cancer Institute and the Margaret A. Cunningham Immune Mechanisms in Cancer Research Fellowship Award (M.D.) from the John D. Proctor Foundation. T.J. is a Howard Hughes Investigator and a Daniel K. Ludwig Scholar.

Author Contributions M.D. and T.J. designed the study. M.D. performed all experiments with assistance from C.M. and L.M.S. A.F.C. provided reagents and conceptual advice. M.D. and T.J. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to T.J. (tjacks@mit.edu).

METHODS

Mice and tumour induction. 129S₄/SvJae strains backcrossed 8 generations were used for all experiments. *Trp53*^{fl} mice were provided by A. Berns, *Kras*^{LSL-G12D} were generated in our laboratory, and *Rag2*^{-/-} mice were purchased from The Jackson Laboratory. Sarcomas were induced in KP and KPR mice by intramuscular injection of the left hind limb with replication-incompetent lentiviruses expressing Cre recombinase as reported previously^{5,6}. Mice were monitored twice weekly for palpable sarcoma formation beginning 50 days after intramuscular injection. All animal studies and procedures were approved by the Massachusetts Institute of Technology's Committee for Animal Care.

Lentiviral production. Lentivirus was produced by transfection of 293T cells with Δ8.2 (gag/pol), CMV-VSV-G, and the various transfer vectors expressing Cre as described²⁷.

Antibody depletion. Anti-CD4 (GK1.5) and anti-CD8 (YTS169.4) antibodies were administered at a dose of 250 μg per mouse by i.p. injection once weekly for the duration of the experiment.

Preparation, culture and transplantation of primary sarcomas. Primary sarcomas were explanted and single cell suspensions were generated by mincing and digesting the tissues for ~1 h at 37 °C in 125 U ml⁻¹ collagenase type I (Gibco), 60 U ml⁻¹ hyaluronidase (Sigma), and 2 mg ml⁻¹ collagenase/dispase (Roche), followed by passage through a 70 μm filter. Subcutaneous transplantation used 2 × 10⁵ cells from freshly isolated tumour cells or cell lines from primary autochthonous tumours that were trypsinized and washed three times in plain DME medium. Transplant recipients were immune-competent or *Rag2*^{-/-} mice on the 129S₄/SvJae background from the same mouse colony used to generate the autochthonous tumours. Subcutaneously transplanted tumour volumes were calculated by multiplying the length × width × height of each tumour. In Fig. 3, the mean volume ± s.e.m. of each tumour line is depicted after transplantation into wild-type mice (WT, open columns) at the time point when the same tumour line reached a volume of 1,000 mm³ in the *Rag2*^{null} transplanted mice (*Rag2*^{null}, filled columns).

Flow cytometry. Cell suspensions from lymphoid organs were prepared by mechanical disruption between frosted slides. Cells were then stained with antibodies for 20–30 min after treatment with FcBlock (BD Pharmingen). Anti-CD8α (53-6.7), anti-IFNγ (XMG1.2), anti-TNFα (MP6-XT22), and DimerX I (Dimeric

Mouse H-2K^b:Ig) were from BD Pharmingen. All antibodies were used at 1:200 dilution. Peptide-loaded DimerX reagents were prepared as directed and used at 1:75 dilution. To improve the sensitivity of the DimerX reagent, we used both PE and APC labelled dimers to co-stain CD8⁺ T cells. Propidium iodide was used to exclude dead cells. Cells were read on a FACSCalibur and analysed using Flowjo software (Tree Star). In Fig. 2c, d, data were determined by comparing the fraction of CD8⁺ cells in duplicate samples stained with K^b dimers or for cytokine production and exceeds 100% due to the incomplete sensitivity of the K^b dimers to detect antigen specific cells. In Fig. 4c, freshly harvested sarcomas were cultured for 24 h in the presence of 10 U IFN-γ (solid line) or untreated (dashed line) and analysed for H-2K^b surface expression (shaded, control antibody).

Cytokine production. Cells were resuspended in the presence or absence of SIYRYGL and SIINFEKL peptides in OPTI-MEM I (Gibco) supplemented with GolgiPlug (BD Pharmingen) for ~4 h at 37 °C, 5% CO₂. Cells were then fixed and stained for intracellular cytokines using the Cytofix/Cytoperm kit (BD Biosciences).

Luciferase detection. Freshly explanted tumours or cell lines were lysed in Cell Culture Lysis Reagent, mixed with Luciferase Assay Reagent according to the manufacturer's instructions (Promega), and relative light units (RLU) were detected using the Optocomp I luminometer (MGM Instruments). RLUs were standardized by the total amount of protein (Bio-Rad Protein Assay) in each sample. *In vivo* bioluminescence images were acquired with the NightOWLII LB983 (Berthold Technologies) or the IVIS Spectrum (Xenogen Corp.) after intraperitoneal injection of 1.5 mg beetle luciferin (Promega).

5-aza-2'-deoxycytidine treatment. Tumour cell lines were plated at low confluency (2 × 10⁵ cells per well of 6-well plate), and treated with 1 μM 5-aza-2'-deoxycytidine replaced daily for three consecutive days and then analysed for luciferase activity.

Influenza. WSN-SIY (20 p.f.u. per mouse) provided by J. Chen. FACs analysis performed four months after intratracheal infection.

Statistical analyses. *P*-values were generated using unpaired two-tailed Fisher exact probability tests or Student's *t*-tests.

27. DuPage, M., Dooley, A. L. & Jacks, T. Conditional mouse lung cancer models using adenoviral or lentiviral delivery of Cre recombinase. *Nature Protocols* **4**, 1064–1072 (2009).

Galectin 8 targets damaged vesicles for autophagy to defend cells against bacterial invasion

Teresa L. M. Thurston¹, Michal P. Wandel¹, Natalia von Muhlinen¹, Ágnes Foeglein¹ & Felix Randow¹

Autophagy defends the mammalian cytosol against bacterial infection^{1–3}. Efficient pathogen engulfment is mediated by cargo-selecting autophagy adaptors that rely on unidentified pattern-recognition or danger receptors to label invading pathogens as autophagy cargo, typically by polyubiquitin coating^{4–9}. Here we show in human cells that galectin 8 (also known as LGALS8), a cytosolic lectin, is a danger receptor that restricts *Salmonella* proliferation. Galectin 8 monitors endosomal and lysosomal integrity and detects bacterial invasion by binding host glycans exposed on damaged *Salmonella*-containing vacuoles. By recruiting NDP52 (also known as CALCOCO2), galectin 8 activates antibacterial autophagy. Galectin-8-dependent recruitment of NDP52 to *Salmonella*-containing vesicles is transient and followed by ubiquitin-dependent NDP52 recruitment. Because galectin 8 also detects sterile damage to endosomes or lysosomes, as well as invasion by *Listeria* or *Shigella*, we suggest that galectin 8 serves as a versatile receptor for vesicle-damaging pathogens. Our results illustrate how cells deploy the danger receptor galectin 8 to combat infection by monitoring endosomal and lysosomal integrity on the basis of the specific lack of complex carbohydrates in the cytosol.

Galectins are β -galactoside-binding lectins that accumulate in the cytosol before being secreted via a leader-peptide-independent pathway^{10,11}. The best-characterized functions of galectins are performed extracellularly, where they bind glycans to modulate cellular behaviour. However, the occurrence of galectins in the cytosol, which under physiological conditions is devoid of complex carbohydrates, makes them prime candidates for a role as danger and/or pattern-recognition receptors. Galectin 3 (also known as LGALS3) accumulates on damaged bacteria-containing vesicles, although the functional consequences of its recruitment remain unknown^{12,13}. We screened a panel of human galectins for their ability to detect invasion by *Salmonella enterica* serovar Typhimurium. At 1 h post-infection (p.i.), galectin 3, 8 and 9 accumulated on about 10% of intracellular *S. Typhimurium* (Fig. 1a, b and Supplementary Fig. 1a), of which 90% were associated with LAMP1 (Supplementary Fig. 1b). Recruitment of galectins peaked between 1 h and 2 h p.i. (Supplementary Fig. 1c). As galectin 3, 8 and 9 were recruited to *Salmonella*-containing vesicles (SCVs), we used short interfering RNAs (siRNAs) to test whether their depletion causes hyperproliferation of *S. Typhimurium*. Cells lacking galectin 8 or NDP52, but not galectin 3 and/or 9, failed to suppress proliferation of *S. Typhimurium* (Fig. 1c and Supplementary Figs 2a–c and 3a). Microscopic analysis confirmed that the greater bacterial burden of cells lacking galectin 8 was caused by enhanced proliferation rather than differential uptake of bacteria (Supplementary Fig. 3b). Hyperproliferating bacteria in cells lacking galectin 8 appeared mainly in a LAMP1-negative compartment (Supplementary Fig. 3c), consistent with colonization of the cytosol. We conclude that galectin 8 is an antibacterial restriction factor.

As autophagy provides antibacterial protection to cells, the decoration of SCVs with galectins might be an autophagy-inducing signal, analogous to ubiquitin coating^{14,15}. We therefore tested binding of galectins to autophagy receptors that restrict the proliferation of *S. Typhimurium*, that is, NDP52, p62 and optineurin^{7–9}. We found in a

luminescence-based mammalian interactome mapping (LUMIER) assay that galectin 8 and NDP52 interacted specifically (Fig. 2a and Supplementary Fig. 4a). Binding was confirmed by precipitating endogenous NDP52 with Flag-tagged galectin 8 (Fig. 2b).

SCVs double labelled by endogenous galectin 8 and NDP52 were prominent in *Salmonella*-infected HeLa cells (Fig. 2c). In cells expressing yellow fluorescent protein (YFP)-tagged galectins the majority of

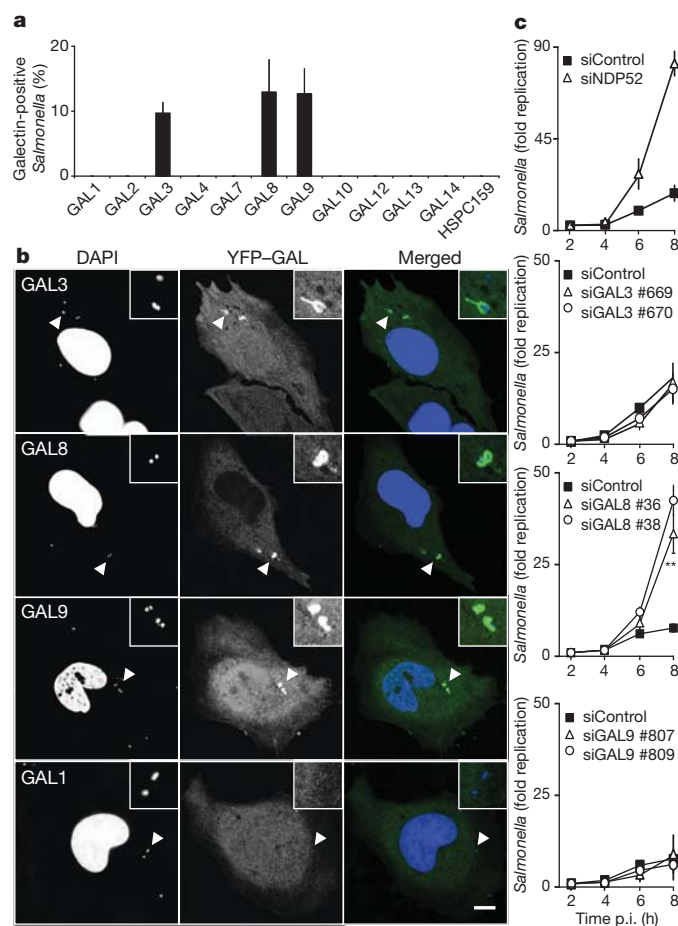


Figure 1 | Galectin 8 responds to infection by *S. Typhimurium* and restricts bacterial proliferation. **a, b**, Analysis of HeLa cells stably expressing YFP fused to the indicated galectins and infected with *S. Typhimurium* for 1 h. **a**, Percentage of bacteria coated by the indicated galectins. YFP-positive bacteria were counted by microscopy. Mean and standard deviation (s.d.) of triplicate HeLa cultures, $n > 100$ bacteria per coverslip. **b**, Confocal micrographs. Arrowheads, bacteria shown in insets. DAPI, 4',6-diamidino-2-phenylindole. **c**, Kinetics of fold replication for *S. Typhimurium* in HeLa cells transfected with the indicated siRNAs. Bacteria were counted on the basis of their ability to form colonies on agar plates. Mean and s.d. of triplicate HeLa cultures and duplicate colony counts. siRNAs are further characterized in Supplementary Fig. 2a–c. $**P < 0.01$, Student's *t*-test. Scale bar, 10 μ m.

¹MRC Laboratory of Molecular Biology, Division of Protein and Nucleic Acid Chemistry, Hills Road, Cambridge CB2 0QH, UK.

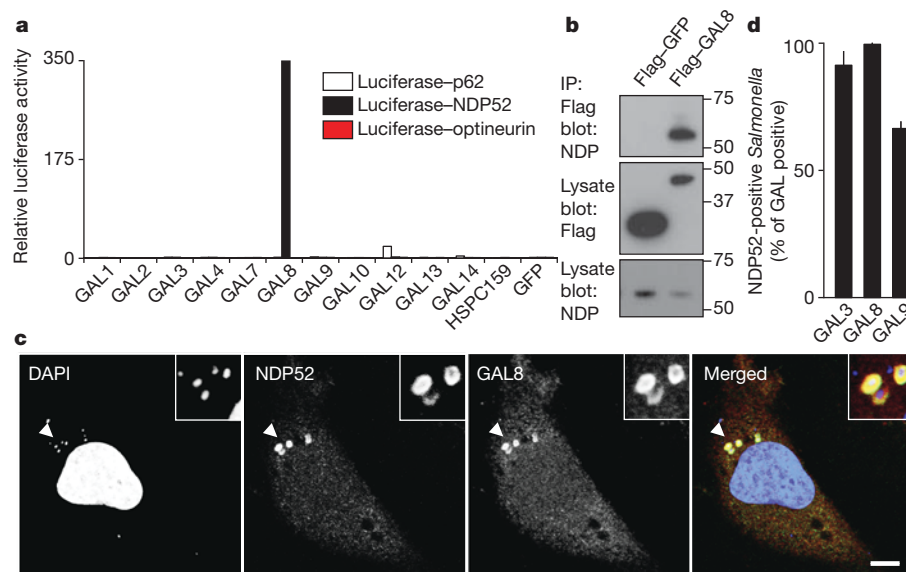


Figure 2 | Galectin 8 binds NDP52. **a**, LUMIER binding assay: normalized ratio between luciferase activity bound to beads and present in lysates. Lysates of 293ET cells expressing NDP52, p62 or optineurin each fused to luciferase, and the indicated Flag-tagged galectins were incubated with anti-Flag beads. Flag-tagged proteins are further characterized in Supplementary Fig. 4a. **b**, Lysates of 293ET cells, expressing Flag-tagged proteins as indicated, were immunoprecipitated with anti-Flag beads. Lysates and immunoprecipitates (IP) were blotted for the presence of Flag-tagged proteins and endogenous

NDP52. **c**, Confocal images of HeLa cells infected with *S. Typhimurium* for 1 h and stained with antisera against NDP52 and galectin 8. Arrowheads, bacteria shown in insets. **d**, Co-localization of NDP52 with galectin-positive bacteria in HeLa cells stably expressing YFP fused to the indicated galectins, infected with *S. Typhimurium* and stained with NDP52 antiserum 1 h after infection. Mean and s.d. of duplicate HeLa cultures, $n > 100$ bacteria per coverslip, representative of two independent experiments. Scale bar, 10 μ m.

galectin-positive SCVs had accumulated NDP52 (Fig. 2d and Supplementary Fig. 5a). Furthermore, at 1 h p.i. NDP52 and galectin 8 co-localized tightly in a pattern distinct from p62 or ubiquitin 'microdomains'¹⁶ (Supplementary Fig. 5b, c).

To characterize further the interaction between galectin 8 and NDP52 we determined their respective binding sites. Galectin 8 contains two carbohydrate-recognition domains (CRD) (Supplementary Fig. 6a). NDP52 bound galectin-8_{A1–228} (that is, amino acids 229–360), equivalent to the second CRD, but not galectin-8_{1–228} (that is, amino acids 1–228) (Supplementary Fig. 6b). NDP52 harbours a SKICH domain, a coiled coils-forming region, and a ubiquitin-binding zinc finger (Supplementary Fig. 6a). Galectin 8 bound NDP52_{1–393} but not NDP52_{1–370} (Supplementary Fig. 6c). The NDP52 fragment spanning residues 370–393 is therefore essential for binding galectin 8. This fragment, as well as NDP52_{372–380}, purified as GST-fusion proteins, bound galectin 8 (Supplementary Fig. 6d). A point mutation within NDP52_{372–380} (L374A) abrogated binding to galectin 8, without compromising binding to ubiquitin when introduced into full-length NDP52 (Supplementary Fig. 6d, e). Binding of galectin 8 to NDP52 is direct, as the purified proteins interacted (Supplementary Fig. 6f).

To determine whether one monomer of the NDP52–galectin-8 heteromeric complex recruits the other partner, the accumulation of galectins on SCVs in cells depleted of NDP52 or TBK1 was analysed (Fig. 3a and Supplementary Fig. 2). Galectin 3, 8 and 9 re-distributed normally to SCVs in siRNA-treated cells. In contrast, in cells depleted of galectin 8 NDP52 did not localize to SCVs at 1 h p.i., a phenotype that was complemented upon expression of siRNA-resistant galectin 8 (Fig. 3b and Supplementary Fig. 7). Cells lacking galectin 3 or galectin 9 had no defect in recruiting NDP52 to SCVs. The recruitment of NDP52 to SCVs is therefore specifically mediated by galectin 8, while NDP52 and TBK1 are dispensable for the accumulation of galectins on SCVs.

Rupture of SCVs exposes the cytosol to host glycans and microbial carbohydrates, either or both of which may cause galectin 8 accumulation at SCVs. The requirement for carbohydrate binding by galectin 8 was tested using point mutations in either CRD¹⁷. In contrast to

galectin-8(R232H), galectin-8(R69H) did not accumulate at SCVs, showing that the amino-terminal CRD is required for carbohydrate-dependent recruitment of galectin 8 to SCVs (Fig. 3c). To test whether the carbohydrates detected by galectin 8 are of microbial origin, binding of recombinant galectin 8 to bacteria *in vitro* was analysed. Galectin 8 did not bind to *S. Typhimurium* but stained blood-group-B-positive bacteria (*Escherichia coli* strain O86)¹⁷ (Fig. 3d), suggesting that galectin 8, when accumulating on SCVs, recognizes host glycans. The occurrence of galectin-8 ligands in host cells was confirmed by staining HeLa cells with recombinant galectin 8 (Fig. 3d). Direct evidence that host glycans recruit galectins to SCVs was obtained from experiments with CHO-Lec3.2.8.1 cells¹⁸, which lack mature glycans and in which recruitment of galectins to SCVs was severely impaired (Fig. 3e). The detection of host glycans on damaged vesicles by galectin 8 suggests that it is not a receptor specific for *S. Typhimurium*. We therefore tested whether sterile damage to vesicles is detected by galectins. Osmotic damage of endosomes induced dense puncta formed by galectin 3, 8 and 9 but not by galectin 1 (Fig. 3f and Supplementary Fig. 8). Damage to lysosomes by glycyl-L-phenylalanine 2-naphthylamide (GPN) treatment resulted in the initial loss of lysotracker staining, followed by the appearance of galectin 3, 8 and 9 speckles (Supplementary Fig. 9a). In contrast to damaged SCVs and endosomes, burst lysosomes were also detected by galectin 1, suggesting compartment-specific differences in the distribution of galectin ligands. GPN failed to induce speckles of galectin-8(R69H) (Supplementary Fig. 9b), thereby indicating that binding of glycans to the N-terminal CRD of galectin 8 is required to detect lysosomal damage. The capacity of galectin 3, 8 and 9 to detect vesicle damage by binding exposed host glycans suggests their ability to sense the invasion of cells by a wide range of vesicle-damaging pathogens. Indeed, galectin 3, 8 and 9 also accumulated around Gram-positive *Listeria monocytogenes* and Gram-negative *Shigella flexneri* (Supplementary Fig. 10), proving that these galectins detect the invasion of cells by phylogenetically distant bacteria. We conclude that galectin 3, 8 and 9 are danger receptors that sense the exposure of host glycans on ruptured membranes and thereby monitor the integrity of the endosomal/lysosomal compartment.

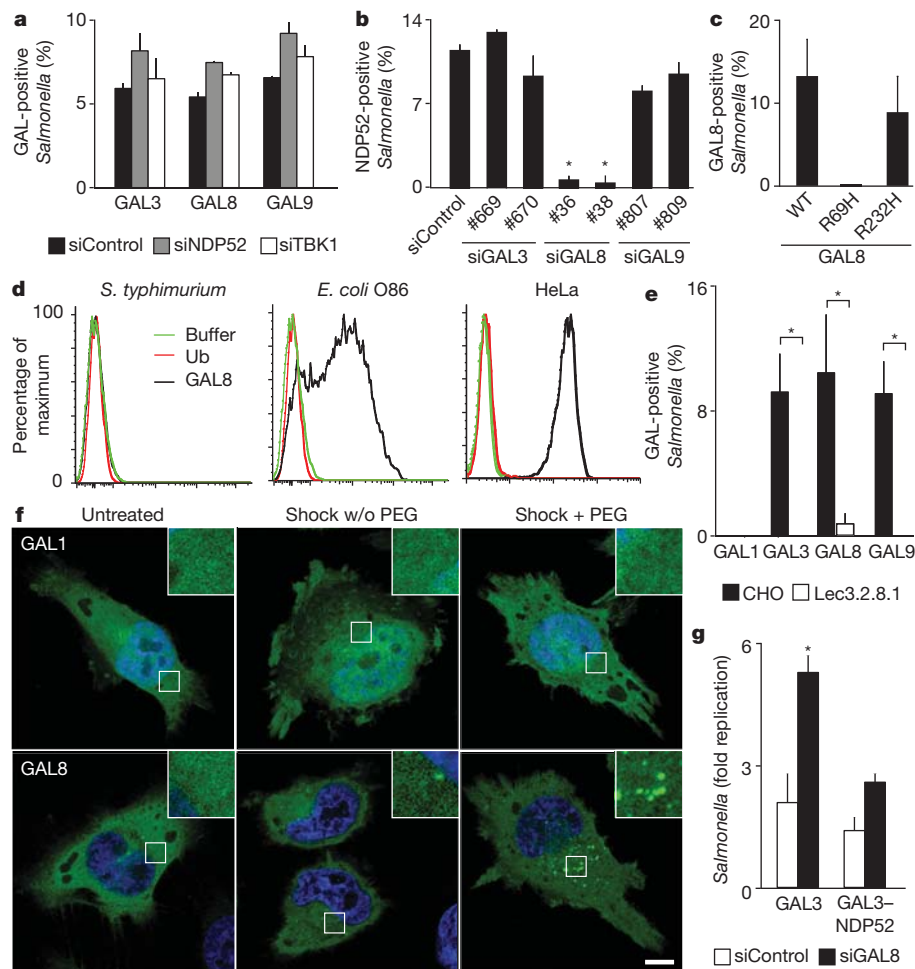


Figure 3 | Galectin 8 is a danger receptor that senses cytosolic host glycans and recruits NDP52 to restrict *Salmonella* proliferation. **a**, Percentage of *S. Typhimurium* coated by the indicated galectins. HeLa cells stably expressing YFP-tagged galectins were treated with the indicated siRNAs. YFP-positive bacteria were counted by microscopy at 1 h p.i. siRNAs are further characterized in Supplementary Fig. 2. **b**, Analysis of HeLa cells treated with the indicated siRNAs and stained with NDP52 antiserum. NDP52-positive bacteria were counted by microscopy 1 h after infection with *S. Typhimurium*. siRNAs are further characterized in Supplementary Fig. 2. **c**, Percentage of bacteria coated by the indicated galectin 8 alleles. HeLa cells stably expressing the indicated galectin 8 alleles fused to YFP were infected with *S. Typhimurium*. YFP-positive *S. Typhimurium* were counted by microscopy at 75 min p.i. WT, wild type. **d**, Binding of galectin 8 to bacteria and HeLa cells. The indicated bacteria and HeLa cells were incubated with His-GST-ubiquitin (Ub), His-GST-galectin 8 or buffer as indicated, followed by murine anti-His antibody

and PE-labelled anti-mouse serum. **e**, Percentage of *S. Typhimurium* coated by the indicated galectins. Wild-type CHO cells and mutant Lec3.2.8.1 cells stably expressing YFP-tagged galectins were infected with *S. Typhimurium*. YFP-positive bacteria were counted by microscopy at 1 h p.i. **f**, Confocal images of HeLa cells expressing the indicated YFP-tagged galectins. Cells were left untreated or were exposed to hypertonic conditions, with or without (w/o) PEG as indicated, followed by hypotonic shock. **g**, Fold replication of *S. Typhimurium* in HeLa cells expressing the indicated galectin 3 variants and transfected with the indicated siRNAs. At 2 h and 6 h after infection, cells were lysed and bacteria counted on the basis of their ability to form colonies on agar plates. Galectin 3 proteins are further characterized in Supplementary Fig. 4b. Mean and s.d. of duplicate coverslips (**a–c**, **e**) or triplicate HeLa cultures and duplicate colony counts (**g**). >100 bacteria counted per coverslip. Data are representative of at least two repeats. * $P < 0.05$, Student's *t*-test. Scale bar, 10 μ m.

To test whether the recruitment of NDP52 to SCVs is essential for the antibacterial function of galectin 8, we depleted cells of galectin 8 and targeted NDP52 artificially to SCVs by fusing it to galectin 3 (Fig. 3g and Supplementary Fig. 4b). *S. Typhimurium* hyperproliferated in galectin-8-depleted cells, while expression of galectin 3 fused to NDP52 restored the cells' restrictive capacity. Artificial targeting of NDP52 to SCVs therefore compensates for the lack of galectin 8, which strongly suggests that the recruitment of NDP52 via galectin 8 is essential to efficiently antagonize bacterial invasion.

NDP52 restricts the proliferation of *S. Typhimurium* by targeting bacteria for autophagy⁸. As galectin 8 recruits NDP52 to SCVs, we investigated whether galectin 8 is required upstream of NDP52 for the induction of antibacterial autophagy. First, we confirmed that at 1 h p.i., *S. Typhimurium* that had been sensed by galectin-8, and therefore had acquired an NDP52 coat, were taken up into LC3-positive autophagosomes (Fig. 4a). Such an outcome was predicted from the

pairwise co-localization of NDP52 with galectin-8- and LC3-positive bacteria (Fig. 2d and Supplementary Fig. 11). We then tested whether depletion of galectin 8 impairs autophagy of *S. Typhimurium*. In the absence of galectin 8 fewer bacteria were targeted by LC3 (Fig. 4b) and of the remaining LC3-positive bacteria fewer had accumulated NDP52 (Supplementary Fig. 11). In contrast, galectin-8 recruitment to SCVs did not require autophagy as it occurred undisturbed in *ATG5*^{-/-} fibroblasts (Supplementary Fig. 12). We conclude that the danger receptor galectin 8, by recruiting NDP52, directs autophagy towards invading bacteria.

The recruitment of NDP52 to invading bacteria is mediated by two signals, the newly discovered carbohydrate-dependent galectin-8 pathway and the previously known ubiquitin-dependent pathway⁸. Their differential contribution to the recruitment of NDP52 to *S. Typhimurium* was investigated by analysing NDP52 mutants selectively disabled to bind galectin 8 and/or ubiquitin. For accurate scoring,

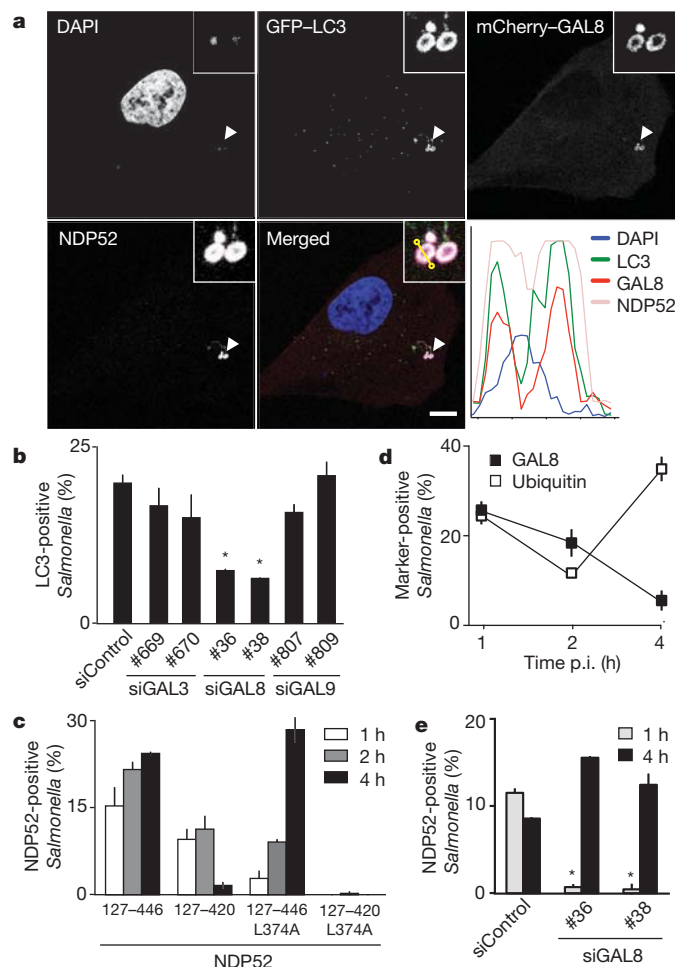


Figure 4 | The antibacterial effect of galectin 8 is mediated by autophagy. **a**, Confocal micrograph of HeLa cells expressing GFP-LC3 and mCherry-galectin-8, stained for NDP52 1 h after infection with *S. Typhimurium*. The lower right panel contains a fluorescence line scan along the yellow line in the merge inset. Arrowheads, bacteria shown in insets. **b**, Percentage of GFP-LC3-positive *S. Typhimurium* at 1 h p.i. in HeLa cells treated with the indicated siRNAs. **c**, Percentage of bacteria positive for NDP52. HeLa cells expressing the indicated NDP52 variants fused to YFP were infected with *S. Typhimurium*. **d**, Percentage of bacteria positive for the indicated markers. HeLa cells, either wild type or expressing YFP-galectin-8 as indicated, were infected with *S. Typhimurium*. Ubiquitin was detected by antibody staining. **e**, HeLa cells treated with the indicated siRNAs were infected with *S. Typhimurium* and stained for NDP52 at the indicated time points. Fluorescent bacteria were counted by microscopy at the indicated time points. Graphs, representing at least two independent repeats, show mean and s.d. of duplicate coverslips for which >200 bacteria were counted. * $P < 0.05$, Student's *t*-test. Scale bar, 10 μ m.

NDP52 Δ KICH was used, as this truncated allele is distributed diffusely throughout the cytosol. NDP52_{127–446} associated with SCVs at all time points investigated (Fig. 4c). Deleting the carboxy-terminal ubiquitin-binding zinc finger (NDP52_{127–420}) impaired the recruitment of NDP52 to *S. Typhimurium* at late but not early time points. In contrast, NDP52_{127–446}(L374A), which lacks affinity for galectin 8 but not ubiquitin (Supplementary Fig. 6e), did not co-localize with bacteria at 1 h p.i. but accumulated progressively over time (Fig. 4c). NDP52 is therefore recruited to SCVs in two phases—an early transient surge driven by galectin 8 and a later wave dependent on ubiquitin. The kinetics of galectin 8 and ubiquitin recruitment to SCVs support this model as SCVs are marked by galectin 8 only at early time points, whereas ubiquitin marks persist (Fig. 4d). Direct evidence for early galectin-8-dependent and late galectin-8-independent recruitment of NDP52 to *S. Typhimurium* was obtained from cells depleted of galectin 8, in which NDP52 and bacteria co-localized at 4 h but not at 1 h p.i.

(Fig. 4e). NDP52_{127–420}(L374A), deficient in binding to ubiquitin and galectin 8, did not translocate to SCVs at any time point tested (Fig. 4c). Taken together, NDP52 relocates to SCVs in response to two signals, which are active against bacteria at different stages of invasion. The early response to invading bacteria requires the galectin-8-dependent pathway, whereas the zinc-finger-dependent pathway dominates at later time points.

The galectin-8/NDP52 pathway sheds light on why most intracellular bacteria avoid the cytosol and prefer vesicular compartments. The cytosol seems to be protected by synergistic layers of antibacterial defence that activate autophagy at distinct steps of the invasion process. An early line of defence comprises the accumulation of diacylglycerol on bacteria-containing vesicles, which subsequently become the target of autophagy¹⁹. Bacteria escaping the diacylglycerol pathway and exposing host glycans on their damaged vacuoles are targeted by galectin 8 and NDP52, as described in this work. A third layer of defence coats invading bacteria with polyubiquitin^{6–9,20}. Neither the enzymatic machinery for nor the substrate of ubiquitylation have been identified, although LRSAM1, a RING-finger E3 ubiquitin ligase, contributes to autophagy of *S. Typhimurium*²¹. Peptidoglycan and septin cages surrounding cytosolic bacteria also contribute to autophagy^{22–25}. Defects in this intricate network of autophagy-inducing defence pathways are likely to cause susceptibility to infection and promote inflammation, for example in Crohn's disease^{26–29}. Galectin 8 is positioned strategically at the cellular entry point for a variety of pathogens and is therefore expected to have shaped pathogen evolution.

METHODS SUMMARY

Galectins were cloned as YFP fusions and transduced into HeLa cells. HeLa cells were infected with *S. Typhimurium* strain 12023. For confocal microscopy cells were fixed in paraformaldehyde. Bacterial growth was assessed by a gentamycin protection assay. Knockdowns were accomplished with Stealth siRNAs. LUMIER assays were performed as described³⁰. For flow-cytometric analysis samples were incubated with lysates of *E. coli* expressing His-GST fusion proteins, followed by anti-His antibody and goat anti-mouse serum. Statistical testing was performed using two-tailed Student's *t*-test.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 28 June; accepted 1 December 2011.

Published online 15 January 2012.

- Yang, Z. & Klionsky, D. J. An overview of the molecular mechanism of autophagy. *Curr. Top. Microbiol. Immunol.* **335**, 1–32 (2009).
- Deretic, V. Autophagy in immunity and cell-autonomous defense against intracellular microbes. *Immunol. Rev.* **240**, 92–104 (2011).
- Levine, B., Mizushima, N. & Virgin, H. W. Autophagy in immunity and inflammation. *Nature* **469**, 323–335 (2011).
- Birmingham, C. L., Smith, A. C., Bakowski, M. A., Yoshimori, T. & Brumell, J. H. Autophagy controls *Salmonella* infection in response to damage to the *Salmonella*-containing vacuole. *J. Biol. Chem.* **281**, 11374–11383 (2006).
- Johansen, T. & Lamark, T. Selective autophagy mediated by autophagic adapter proteins. *Autophagy* **7**, 279–296 (2011).
- Perrin, A., Jiang, X., Birmingham, C., So, N. & Brumell, J. Recognition of bacteria in the cytosol of mammalian cells by the ubiquitin system. *Curr. Biol.* **14**, 806–811 (2004).
- Zheng, Y. T. et al. The adaptor protein p62/SQSTM1 targets invading bacteria to the autophagy pathway. *J. Immunol.* **183**, 5909–5916 (2009).
- Thurston, T. L. M., Ryzhakov, G., Bloor, S., von Muhlen, N. & Randow, F. The TBK1 adaptor and autophagy receptor NDP52 restricts the proliferation of ubiquitin-coated bacteria. *Nature Immunol.* **10**, 1215–1221 (2009).
- Wild, P. et al. Phosphorylation of the autophagy receptor optineurin restricts *Salmonella* growth. *Science* **333**, 228–233 (2011).
- Houzelstein, D. et al. Phylogenetic analysis of the vertebrate galectin family. *Mol. Biol. Evol.* **21**, 1177–1187 (2004).
- Rabinovich, G. A. & Toscano, M. A. Turning “sweet” on immunity: galectin-glycan interactions in immune tolerance and inflammation. *Nature Rev. Immunol.* **9**, 338–352 (2009).
- Paz, I. et al. Galectin-3, a marker for vacuole lysis by invasive pathogens. *Cell. Microbiol.* **12**, 530–544 (2009).
- Dupont, N. et al. *Shigella* phagocytic vacuolar membrane remnants participate in the cellular response to pathogen invasion and are regulated by autophagy. *Cell Host Microbe* **6**, 137–149 (2009).
- Randow, F. How cells deploy ubiquitin and autophagy to defend their cytosol from bacterial invasion. *Autophagy* **7**, 304–309 (2011).

15. Shahnazari, S. & Brumell, J. H. Mechanisms and consequences of bacterial targeting by the autophagy pathway. *Curr. Opin. Microbiol.* **14**, 68–75 (2011).
16. Cemma, M., Kim, P. K. & Brumell, J. H. The ubiquitin-binding adaptor proteins p62/SQSTM1 and NDP52 are recruited independently to bacteria-associated microdomains to target *Salmonella* to the autophagy pathway. *Autophagy* **7**, 341–345 (2011).
17. Stowell, S. R. *et al.* Innate immune lectins kill bacteria expressing blood group antigen. *Nature Med.* **16**, 295–301 (2010).
18. Patnaik, S. K. & Stanley, P. Lectin-resistant CHO glycosylation mutants. *Methods Enzymol.* **416**, 159–182 (2006).
19. Shahnazari, S. *et al.* A diacylglycerol-dependent signaling pathway contributes to regulation of antibacterial autophagy. *Cell Host Microbe* **8**, 137–146 (2010).
20. Collins, C. A. *et al.* Atg5-independent sequestration of ubiquitinated mycobacteria. *PLoS Pathog.* **5**, e1000430 (2009).
21. Ng, A. C. Y. *et al.* Human leucine-rich repeat proteins: a genome-wide bioinformatic categorization and functional analysis in innate immunity. *Proc. Natl Acad. Sci. USA* **108**, 4631–4638 (2011).
22. Mostowy, S. *et al.* Entrapment of intracytosolic bacteria by septin cage-like structures. *Cell Host Microbe* **8**, 433–444 (2010).
23. Yano, T. *et al.* Autophagic control of *Listeria* through intracellular innate immune recognition in *Drosophila*. *Nature Immunol.* **9**, 908–916 (2008).
24. Travassos, L. H. *et al.* Nod1 and Nod2 direct autophagy by recruiting ATG16L1 to the plasma membrane at the site of bacterial entry. *Nature Immunol.* **11**, 55–62 (2010).
25. Cooney, R. *et al.* NOD2 stimulation induces autophagy in dendritic cells influencing bacterial handling and antigen presentation. *Nature Med.* **16**, 90–97 (2010).
26. Hugot, J. P. *et al.* Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* **411**, 599–603 (2001).
27. McCarroll, S. A. *et al.* Deletion polymorphism upstream of *IRGM* associated with altered *IRGM* expression and Crohn's disease. *Nature Genet.* **40**, 1107–1112 (2008).
28. Rioux, J. D. *et al.* Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. *Nature Genet.* **39**, 596–604 (2007).
29. Hampe, J. *et al.* A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in *ATG16L1*. *Nature Genet.* **39**, 207–211 (2007).
30. Ryzhakov, G. & Randow, F. SINTBAD, a novel component of innate antiviral immunity, shares a TBK1-binding domain with NAPI and TANK. *EMBO J.* **26**, 3180–3190 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Kendrick-Jones (MRC Laboratory of Molecular Biology), A. Geerlof (European Molecular Biology Laboratory Heidelberg), N. Mizushima (Tokyo University) and P. Stanley (Albert Einstein College of Medicine) for kindly sharing reagents.

Author Contributions T.L.M.T., M.P.W., N.v.M., Á.F. and F.R. planned, performed and analysed experiments. T.L.M.T. and F.R. designed the overall research. F.R. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to F.R. (randow@mrc-lmb.cam.ac.uk).

METHODS

Antibodies. Antibodies were from QIAGEN (Penta-His), the Developmental Studies Hybridoma Bank (LAMP1), BD Transduction Laboratories (p62), Santa Cruz (GAL8-H80, TBK1-C100), R&D Systems (galectin 8), Transduction Laboratories (NDP52, for western blots), Enzo Life Science (ubiquitin FK2), Sigma (ATG5, Flag M2), Dabco (HRP-conjugated reagents), Jackson ImmunoResearch Laboratories (goat anti-mouse-phycoerythrin (PE)) and Invitrogen (Alexa-conjugated anti-mouse and anti-rabbit antisera). The antiserum against NDP52 used for immunofluorescence was a gift from J. Kendrick-Jones.

Plasmids. M5P or closely related plasmids were used to produce recombinant MLV for the expression of proteins in mammalian cells³¹. pETM plasmids were gifts from A. Geerloff. Open reading frames encoding human galectins, NDP52, p62, optineurin, ubiquitin, ATG5 and LC3C were amplified by PCR or have been described^{8,32}. Mutations were generated by PCR and verified by sequencing.

Bacteria. *S. Typhimurium* (strain 12023), provided by D. Holden, was grown overnight in Luria broth (LB) and sub-cultured (1:33) in fresh LB for 3.5 h before infection. HeLa cells in 24-well plates were infected with 20 μ l of such cultures for 15 min at 37 °C. Following two washes with warm PBS and an incubation with 100 μ g ml⁻¹ gentamycin for 2 h cells were cultured in 20 μ g ml⁻¹ gentamycin. To enumerate intracellular bacteria, cells from triplicate wells were lysed in 1 ml cold PBS containing 0.1% Triton-X-100. Serial dilutions were plated in duplicate on TYE agar.

S. flexneri M90T, provided by C. Tang, was grown overnight in Tryptic Soy Broth (TSB) and sub-cultured (1:100) in fresh TSB for 2 h before infection. Bacteria were resuspended in warm IMDM and 100 μ l were added to HeLa cells in 24-well plates. Samples were centrifuged for 10 min at 670g. Following incubation at 37 °C for 30 min, cells were washed with warm PBS and cultured in 100 μ g ml⁻¹ gentamycin for 2 h and 20 μ g ml⁻¹ thereafter.

L. monocytogenes strain EGD (BUG 600), provided by P. Cossart, was grown overnight in Brain Heart Infusion (BHI) at 30 °C with shaking. Five-hundred microlitres of diluted cultures (1:333) were added to HeLa cells in 24-well plates, which were centrifuged at 670g for 10 min. Cells were incubated for 1 h at 37 °C, washed with warm PBS and cultured in media supplemented with 100 μ g ml⁻¹ gentamycin for the next hour and 20 μ g ml⁻¹ gentamycin thereafter.

Cell culture. Cells were grown in IMDM supplemented with 10% FCS at 37 °C in 5% CO₂. HeLa cells were obtained from the European Collection of Cell Cultures, CHO and Lec3.2.8.1 cells¹⁸ were obtained from P. Stanley, ATG5^{-/-} MEFs³³ from N. Mizushima.

RNA interference. 5 × 10⁴ cells per well were seeded in 24-well plates. The following day, cells were transfected with 40 pmol of siRNA (Invitrogen) using Lipofectamine 2000 (Invitrogen) in OptiMem medium (Invitrogen). OptiMem was replaced with complete IMDM medium after 4 h and experiments were performed after 3 days. siRNAs targeted the following sequences: siNDP52 5'-UUCAGUUGAAGCAGCUCUGUCUCC³; siGAL8 #36 5'-CCCACGCCUGAAUUAUAAAGCAUUU; siGAL8 #38 5'-GGACAAAUUCCAGGUGGCGUAAA; siGAL3 #669 5'-AAGCCCAAUGCAAACAGAAUUGCUU; siGAL3 #670 5'-GAGAACAACAGGAGAGUCAUUGUUU; siGAL9 #807 5'-GGCUUCAGUGGAAAUGACAUGCCU; siGAL9 #809 5'-UGUGCAACACGAGGCGAGAACGGAGG; siTBK1 5'-GACAGAAGUUGAUCACA(TT)³⁴.

To render galectin 8 resistant to siGAL8 #38, silent mutations (underlined) GGATAAGTTTCAAGTGCAGTTAAT were introduced by PCR and confirmed by sequencing.

Immunoprecipitation and western blot. Post-nuclear supernatants from 2 × 10⁶ HeLa cells expressing Flag-tagged proteins were obtained following lysis

(150 mM NaCl, 0.1% Triton-X-100, 20 mM Tris-HCl (pH 7.4), 5 mM EDTA and proteinase inhibitors). Protein complexes were immunoprecipitated for 2 h with Flag agarose before washing. Samples were eluted with Flag peptide and separated on 4–12% denaturing Bis-Tris gels (Invitrogen). Visualization following immunoblotting was performed using ECL detection reagents (Amersham Bioscience).

LUMIER assays. LUMIER binding assays^{30,35} with pairs of putative interactors, one fused to luciferase and the other fused to GST or Flag, were performed in LUMIER lysis buffer (150 mM NaCl, 0.1% Triton-X-100, 20 mM Tris-HCl (pH 7.4), 5% glycerol, 5 mM EDTA and proteinase inhibitors). GST-fusion proteins were immobilized on beads before incubation with the luciferase tagged binding partner for 2 h. For Flag-based assays, both proteins were expressed in 293ET cells and immobilized using Flag-agarose. After washing in lysis buffer, proteins were eluted with glutathione or Flag peptide in *Renilla* lysis buffer (Promega). Relative luciferase activity represents the ratio of activity eluted from beads and present in lysates.

FACS. To examine the binding of galectin 8, bacteria in stationary phase or HeLa cells were washed in PBSF (PBS, 2% FCS) and incubated for 30 min at 4 °C with cleared lysates of *E. coli* expressing His-GST fusion proteins, followed by incubations with anti-His antibody and PE-conjugated goat anti-mouse serum. Bacteria were fixed in 4% paraformaldehyde before analysis.

Sterile damage to vesicles. Endosomes were lysed by exposing cells for 10 min to hypertonic medium (0.5 M sucrose in PBS, with or without 10% PEG1000), followed by two PBS washes and an incubation in 60% PBS for 3 min³⁶. Cells were returned to complete medium for 20 min, before being fixed in paraformaldehyde. For live imaging of lysosomal damage, cells were labelled for 1 h with 100 nM LysoTracker Red (Invitrogen), washed with PBS, incubated in Leibovitz L15 medium and, after acquisition of the first image, exposed to 333 μ M GPN³⁷.

Microscopy. HeLa cells were grown on glass cover slips before infection. After infection, cells were washed twice with warm PBS and fixed in 4% paraformaldehyde in PBS for 30 min. Cells were washed twice in PBS and then quenched with PBS pH 7.4 containing 1 M glycine and 0.1% Triton-X-100 for 30 min before blocking for 30 min in PBTB (PBS, 0.1% Triton-X-100, 2% BSA). Cover slips were incubated with primary followed by secondary antibodies for 1 h in PBTB before being mounting in medium containing DAPI (Vector Laboratories). At least 100 events per slide were scored in quantitative assays. Confocal images were taken with a ×63, 1.4 numerical aperture objective on either a Zeiss 710 or a Zeiss 780 microscope. Live imaging was performed on a Nikon Eclipse Ti equipped with an Andor Revolution XD system and a Yokogawa CSU-X1 spinning disk unit.

1. Randow, F. & Sale, J. E. Retroviral transduction of DT40. *Subcell. Biochem.* **40**, 383–386 (2006).
2. Bloor, S. *et al.* Signal processing by its coil zipper domain activates IKK γ . *Proc. Natl Acad. Sci. USA* **105**, 1279–1284 (2008).
3. Kuma, A. *et al.* The role of autophagy during the early neonatal starvation period. *Nature* **432**, 1032–1036 (2004).
4. Fitzgerald, K. A. *et al.* IKK ϵ and TBK1 are essential components of the IRF3 signaling pathway. *Nature Immunol.* **4**, 491–496 (2003).
5. Barrios-Rodiles, M. *et al.* High-throughput mapping of a dynamic signaling network in mammalian cells. *Science* **307**, 1621–1625 (2005).
6. Shaughnessy, L. M., Lipp, P., Lee, K.-D. & Swanson, J. A. Localization of protein kinase C ϵ to macrophage vacuoles perforated by *Listeria monocytogenes* cytolysin. *Cell. Microbiol.* **9**, 1695–1704 (2007).
7. Berg, T. O., Strømhaug, P. E., Berg, T. & Seglen, P. O. Separation of lysosomes and autophagosomes by means of glycyl-phenylalanine-naphthylamide, a lysosome-disrupting cathepsin-C substrate. *Eur. J. Biochem.* **221**, 595–602 (1994).

Brassinosteroid regulates stomatal development by GSK3-mediated inhibition of a MAPK pathway

Tae-Wuk Kim^{1,2}, Marta Michniewicz³, Dominique C. Bergmann³ & Zhi-Yong Wang¹

Plants must coordinate the regulation of biochemistry and anatomy to optimize photosynthesis and water-use efficiency. The formation of stomata, epidermal pores that facilitate gas exchange, is highly coordinated with other aspects of photosynthetic development. The signalling pathways controlling stomata development are not fully understood^{1,2}, although mitogen-activated protein kinase (MAPK) signalling is known to have key roles. Here we demonstrate in *Arabidopsis* that brassinosteroid regulates stomatal development by activating the MAPK kinase kinase (MAPKKK) YDA (also known as YODA). Genetic analyses indicate that receptor kinase-mediated brassinosteroid signalling inhibits stomatal development through the glycogen synthase kinase 3 (GSK3)-like kinase BIN2, and BIN2 acts upstream of YDA but downstream of the ERECTA family of receptor kinases. Complementary *in vitro* and *in vivo* assays show that BIN2 phosphorylates YDA to inhibit YDA phosphorylation of its substrate MKK4, and that activities of downstream MAPKs are reduced in brassinosteroid-deficient mutants but increased by treatment with either brassinosteroid or GSK3-kinase inhibitor. Our results indicate that brassinosteroid inhibits stomatal development by alleviating GSK3-mediated inhibition of this MAPK module, providing two key links; that of a plant MAPKKK to its upstream regulators and of brassinosteroid to a specific developmental output.

In animals and plants, steroid hormones have important roles in coordinating development and metabolism³. In contrast to animal steroid hormones, which act through nuclear receptor transcription factors³, the plant steroid hormone brassinosteroid binds to the extracellular domain of the membrane-bound receptor kinase brassinosteroid insensitive 1 (BRI1). This activates intracellular signal transduction mediated by the serine/threonine protein kinase BSK1, the protein phosphatase BSU1, the GSK3-like BIN2 kinase, PP2A phosphatase and BRASSINAZOLE RESISTANT 1 (BZR1) family transcription factors^{4–10}. When brassinosteroid levels are low, BZR1 is inactivated owing to phosphorylation by BIN2 (refs 11, 12). Brassinosteroid signalling leads to inactivation of BIN2, and PP2A-mediated dephosphorylation and activation of BZR1 (refs 4, 9, 10) (Supplementary Fig. 1a). Although the brassinosteroid signalling pathway has been characterized, its connections to other signalling and developmental pathways are not fully understood.

Stomata are epidermal pores that control gas exchange between the plant and the atmosphere and are critical for maintaining photosynthetic and water-use efficiency in the plant. The density and distribution of stomata in the epidermis of aerial organs is modulated by intrinsic developmental programs, by hormones and by environmental factors such as light, humidity and carbon dioxide^{1,2,13,14}. The genetically defined signalling pathway that regulates stomatal development includes peptide ligands, a receptor protein (TMM), the ERECTA family of receptor-like kinases (ER, ERL1 and ERL2) and a MAPK module comprised of the MAPK kinase kinase (MAPKKK) YDA, the MAPK kinases (MAPKKs) MKK4, MKK5, MKK7 and MKK9, and MAPKs MPK3 and MPK6 (ref 15). Potential downstream

targets include basic helix–loop–helix (bHLH) transcription factors SPEECHLESS (SPCH), MUTE, FAMA, ICE1 (also known as SCRM) and SCRM2, with SPCH being negatively regulated by direct MPK3- and MPK6-mediated phosphorylation^{16,17} (Supplementary Fig. 1b). It is possible that the MAPK pathway integrates environmental and hormonal inputs to optimize stomatal production, but nothing is known about the nature of these signals and their biochemical mechanisms of MAPK pathway regulation.

Excess stomata have been observed in some brassinosteroid-deficient mutants¹⁸. To elucidate the function of brassinosteroid in regulating stomatal development, we examined the distribution of stomata on leaves of brassinosteroid-deficient and brassinosteroid-signalling mutants. In wild-type *Arabidopsis*, stomata are always distributed with at least one pavement cell between them (Fig. 1a). Brassinosteroid deficiency causes stomatal clusters (Fig. 1b, c), whereas treatment with brassinolide (the most active form of brassinosteroid) reduces stomatal density (Fig. 1d), indicating that brassinosteroid represses stomatal development. The brassinosteroid-insensitive mutants *brl1-116*, quadruple *amiRNA-BSL2,3 bsu1 bsl1 (bsu-q)*⁹, dominant *bin2-1* and plants that overexpress *BIN2* also exhibit stomatal clustering (Fig. 1e–h), and overproduce stomatal precursors (meristemoids and guard mother cells) (Fig. 1u and Supplementary Fig. 2). In contrast to the weak stomatal clustering phenotype of the *det2-1* and *brl1-116* mutants, *bsu-q* showed large stomatal clusters on hypocotyls (Supplementary Fig. 4) and cotyledon surfaces consisting almost entirely of stomata (Fig. 1f, u, and Supplementary Figs 2 and 3). Surprisingly, the hyperactive *bzr1-1D* mutation^{10,19} did not affect stomatal development or suppress the stomatal phenotypes of *brl1-116*, *bsu-q* and *bin2-1*, although it suppressed their dwarf phenotypes (Fig. 1i–n and Supplementary Fig. 5). These results indicate that brassinosteroid regulation of stomatal development is mediated by upstream signalling components that include BRI1, BSU1 and BIN2, but that it is independent of the BIN2 substrate BZR1.

Consistent with increased stomatal development in brassinosteroid-insensitive mutants, fewer stomata were observed in cotyledons of plants overexpressing some of the positive brassinosteroid-signalling components of the BSU1 family (Fig. 1q, u and Supplementary Fig. 6) and in *bin2-3 bil1 bil2* loss-of-function mutants lacking 3/7 brassinosteroid-signalling GSK3-like kinases (Fig. 1o, p, u and Supplementary Fig. 2). We used bikinin (4-[(5-bromopyridin-2-yl)amino]-4-oxobutanoic acid, ChemBridge Corporation), a highly specific inhibitor for the 7 *Arabidopsis* GSK3-like kinases that appear to be involved in brassinosteroid signalling^{9,20,21}, to investigate further the function of brassinosteroid-related GSK3-like kinases in stomatal development. When added to the growth medium, bikinin decreased stomatal production in wild-type plants, fully suppressed the stomatal clustering phenotypes of *bin2-1* and partially suppressed the severe stomatal phenotypes of *bsu-q* (Fig. 1r–u). These results confirm that increased activity of the GSK3-like kinases is responsible for enhanced stomatal production in brassinosteroid-deficient and brassinosteroid-insensitive mutants.

¹Department of Plant Biology, Carnegie Institution for Science, Stanford, California 94305-4150, USA. ²Department of Life Science, Hanyang University, Seoul 133-791, South Korea. ³Department of Biology, Stanford University, Stanford, California 94305-5020, USA.

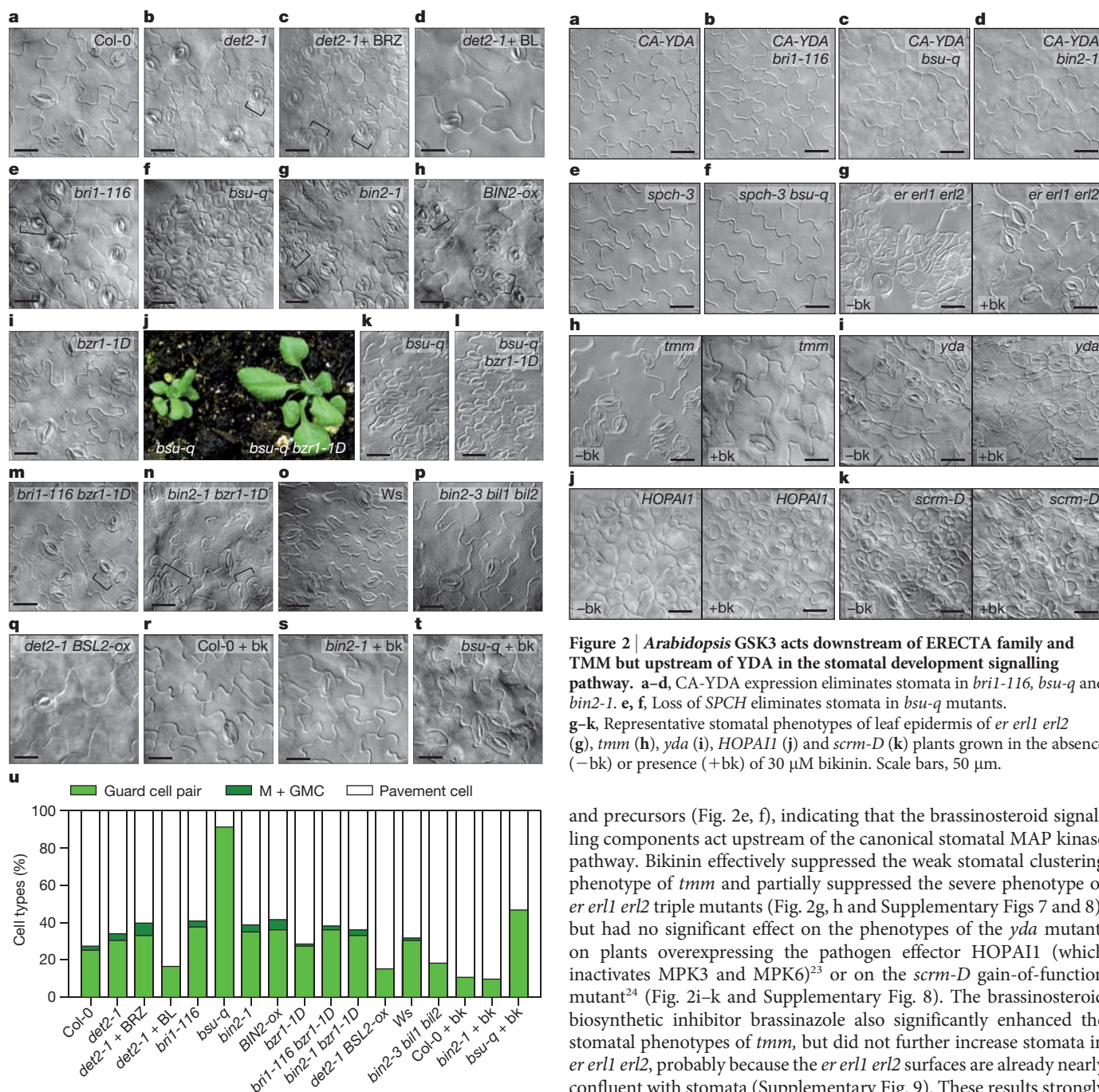


Figure 1 | Brassinosteroid negatively regulates stomatal development. **a–i, k–t**, Differential interference contrast (DIC) microscopy images of abaxial cotyledon epidermis of 8-day-old seedlings or leaf epidermis of 4-week-old plants (**k, l**) with indicated genotypes (Col-0 and Ws are wild-type controls), grown on medium \pm BRZ (2 μ M), brassinolide (BL, 50 nM), or bikinin (bk, 30 μ M). **j**, Growth phenotype of 4-week-old *bsu-q* and *bsu-q bzi1-1D* mutants. **u**, Quantification of epidermal cell types of the indicated 8-day-old mutants, expressed as percentage of total cells. GMC, guard mother cell; M, meristemoid. Brackets in **b, c, e, g, h, m, n** indicate clustered stomata. Scale bars, 50 μ m.

We examined genetic interactions between brassinosteroid mutants and known stomatal mutants. Expression of constitutively active YDA (CA-YDA) can completely eliminate stomatal development²² (Fig. 2a), probably through activation of a MAP kinase pathway that phosphorylates and inactivates SPCH^{15,16}. Expression of CA-YDA completely suppressed stomatal development of the *bri1-116*, *bsu-q* and *bin2-1* mutants (Fig. 2b–d). Loss of SPCH was also completely epistatic to *bsu-q* in that a *bsu-q spch-3* (null) mutant lacked stomata

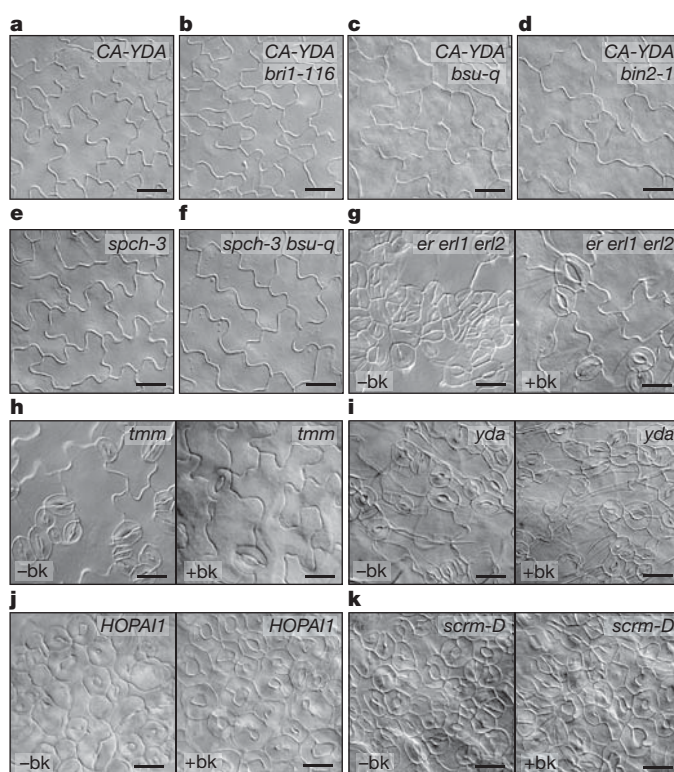


Figure 2 | Arabidopsis GSK3 acts downstream of ERECTA family and TMM but upstream of YDA in the stomatal development signalling pathway. **a–d**, CA-YDA expression eliminates stomata in *bri1-116*, *bsu-q* and *bin2-1*. **e, f**, Loss of SPCH eliminates stomata in *bsu-q* mutants. **g–k**, Representative stomatal phenotypes of leaf epidermis of *er1 erl2* (**g**), *tmm* (**h**), *yda* (**i**), *HOPAI1* (**j**) and *scr-D* (**k**) plants grown in the absence (–bk) or presence (+bk) of 30 μ M bikinin. Scale bars, 50 μ m.

and precursors (Fig. 2e, f), indicating that the brassinosteroid signalling components act upstream of the canonical stomatal MAP kinase pathway. Bikinin effectively suppressed the weak stomatal clustering phenotype of *tmm* and partially suppressed the severe phenotype of *er1 erl2* triple mutants (Fig. 2g, h and Supplementary Figs 7 and 8), but had no significant effect on the phenotypes of the *yda* mutant, on plants overexpressing the pathogen effector HOPAI1 (which inactivates MPK3 and MPK6)²³ or on the *scr-D* gain-of-function mutant²⁴ (Fig. 2i–k and Supplementary Fig. 8). The brassinosteroid biosynthetic inhibitor brassinazole also significantly enhanced the stomatal phenotypes of *tmm*, but did not further increase stomata in *er1 erl2*, probably because the *er1 erl2* surfaces are already nearly confluent with stomata (Supplementary Fig. 9). These results strongly indicate that GSK3-like kinases act downstream of the ER and TMM receptors, but upstream of the YDA MAPKKK.

YDA contains 84 putative GSK3 phosphorylation sites (Ser/Thr-X-X-X-Ser/Thr). Many of these sites are conserved in the two rice homologues of YDA, Os02g0666300 and Os04g0559800, and these homologues also share a highly conserved sequence just amino-terminal of the kinase domain. Importantly, YDA can be made constitutively active when part of this region (amino acids 185–322; Fig. 3a) is deleted²². The region that is deleted in CA-YDA contains 23 putative GSK3 phosphorylation sites, including successive phosphorylation sites that are similar to sites found in the known BIN2 target BZR1 (Fig. 3a and Supplementary Fig. 10).

We tested whether BIN2 directly interacts with and phosphorylates YDA. Maltose binding protein (MBP)-YDA was detected in an overlay assay by using GST-BIN2 and anti-GST antibody (Fig. 3b), demonstrating direct YDA binding to BIN2 *in vitro*. BIN2 also interacted with YDA and CA-YDA in yeast two-hybrid assays (Fig. 3c). *In vitro* kinase

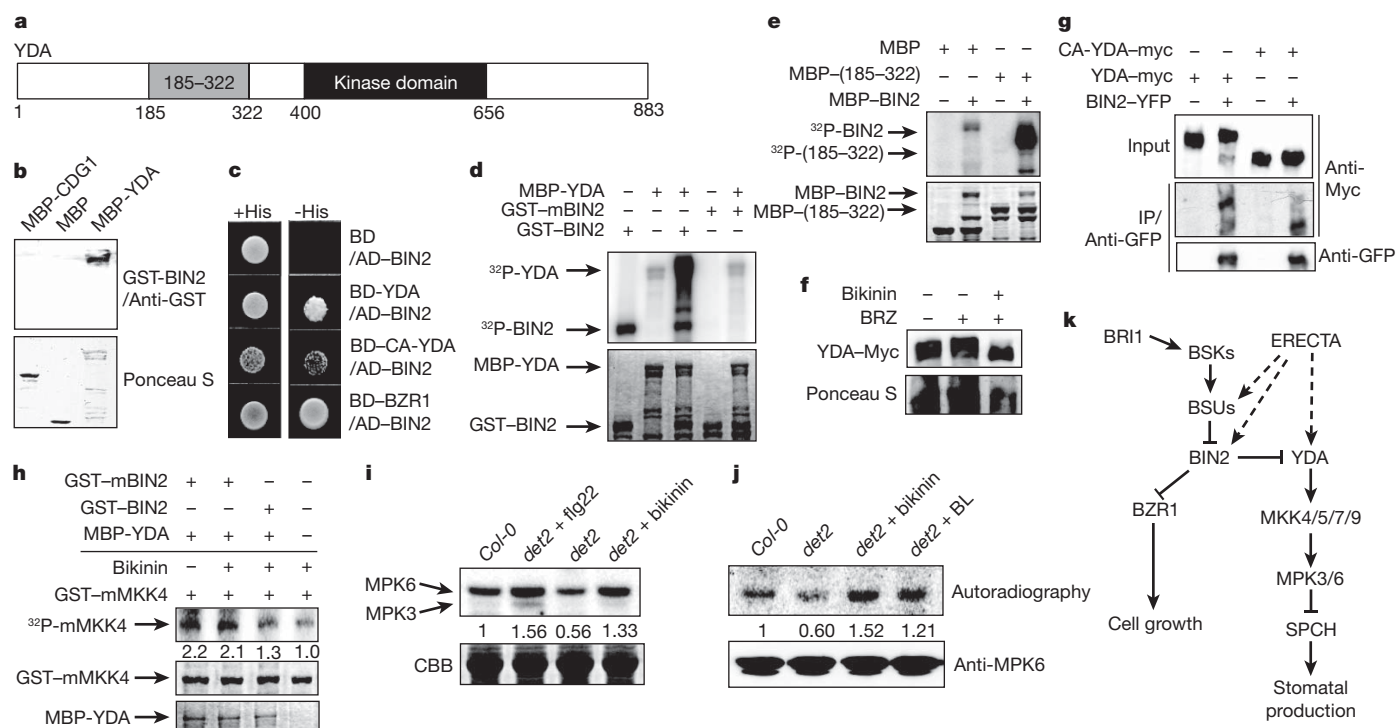


Figure 3 | BIN2 inhibits YDA kinase activity through phosphorylation.

a, Domain structure of YDA. **b**, Gel blot of indicated proteins (MBP-CDG1 is a negative control) sequentially probed with GST-BIN2 and anti-GST-HRP antibody. **c**, Yeast two-hybrid assays of indicated proteins. **d**, **e**, *In vitro* kinase assays of BIN2 phosphorylation of YDA or YDA fragment containing amino acids 185–322 (185–322). Upper panel shows autoradiography and bottom panel shows protein staining. Mutant BIN2 (mBIN2) is kinase inactive. **f**, YDA-Myc plants grown for 5 days on medium containing 2 μ M BRZ \pm 30 μ M bikinin and analysed by anti-Myc immunoblot. **g**, Proteins transiently expressed in *N. benthamiana* leaves, immunoprecipitated (IP) with anti-YFP antibody, and immunoblotted with anti-Myc or anti-YFP antibody. **h**, YDA pre-incubated with BIN2 or mBIN2 (kinase-inactive mutant) and ATP was purified then incubated with mutant MKK4 (mMKK4) and [32 P] γ ATP,

assays showed that BIN2 phosphorylated YDA, but YDA did not phosphorylate a kinase-inactive BIN2 mutant or other brassinosteroid signalling components (Fig. 3d and Supplementary Fig. 11). BIN2 strongly phosphorylated the region deleted in CA-YDA (Fig. 3e), indicating that BIN2 might inhibit YDA by phosphorylating its autoregulatory domain.

BIN2 phosphorylation of BZR1 causes mobility shifts of the phosphorylated BZR1 band in SDS–polyacrylamide gel electrophoresis (SDS–PAGE) gels¹². Like BZR1, YDA that was phosphorylated by BIN2 *in vitro* also exhibited slower mobility (Fig. 3d and Supplementary Fig. 11). Consistent with the *in vitro* data, bikinin treatment of *Arabidopsis* seedlings increased the mobility of YDA-Myc in SDS–PAGE (Fig. 3f). When transiently expressed in *Nicotiana benthamiana* leaf cells, both YDA-Myc and CA-YDA-Myc were co-immunoprecipitated by anti-yellow fluorescent protein (YFP) antibody when co-expressed with BIN2-YFP but not when expressed alone (Fig. 3g), demonstrating that there is an interaction between BIN2 and YDA *in vivo*. Furthermore, co-expression of BIN2 retarded mobility of YDA, but not of CA-YDA bands in immunoblots (Fig. 3g). These results confirm that BIN2 mainly phosphorylates the YDA N-terminal regulatory domain.

Finally, we tested whether BIN2 phosphorylation of YDA affects YDA kinase activity and whether brassinosteroid and bikinin affect MAPK activity in plants. YDA was pre-incubated with BIN2 and ATP, or with a kinase-inactive mutant BIN2 as a control, and then purified and further incubated with MKK4 (its known substrate), bikinin and [32 P] γ ATP. Pre-incubation with BIN2, but not with mutant BIN2,

\pm bikinin. Numbers indicate relative signal levels normalized to loading control. **i–j**, MPK6 and MPK3 activities in seedlings treated with flg22 (10 nM, positive control), bikinin (30 μ M) or BL (100 nM) for 30 min (**i**) or 2 h (**j**), analysed by in-gel kinase assays. Numbers indicate relative signal levels (upper panel) normalized to the loading control (CBB or MPK6 immunoblot). **k**, A model for regulation of stomatal development by two receptor kinase-mediated signal transduction pathways. When brassinosteroid levels are low, BIN2 phosphorylates and inactivates YDA, increasing stomatal production. Brassinosteroid signalling through BRI1 inactivates BIN2, leading to activation of YDA and downstream MAPK proteins, and suppression of stomatal development. ERECTA is genetically upstream of YDA; a biochemical link is not known, but BSU1 and BIN2 or their homologues are strong candidates for intermediates (dashed line).

decreased YDA phosphorylation of MKK4 (Fig. 3h and Supplementary Fig. 12), indicating that BIN2 phosphorylation inhibits YDA activity. Consistent with BIN2 inactivation of YDA, the kinase activities of MPK3 and MPK6 were reduced in the *det2* mutant but increased by treatment with bikinin or brassinolide (Fig. 3i and 3j).

Taken together, our genetic and biochemical analyses demonstrate that brassinosteroid negatively regulates stomatal development by inhibiting the BIN2-mediated phosphorylation and inactivation of YDA (Fig. 3k). When brassinosteroid levels are low, active BIN2 directly phosphorylates and inactivates YDA; reduced MAP kinase pathway activity can de-repress SPCH, allowing SPCH to initiate stomatal development. Brassinosteroid signalling through BRI1, BSK1 and BSU1 inactivates GSK3, resulting in activation of the MAP kinase pathway and inhibition of stomatal production (Fig. 3k).

This study supports a role of brassinosteroid as a master regulator that coordinates both physiological and developmental aspects of plant growth. Previous studies have demonstrated key functions of brassinosteroid in inhibiting photomorphogenesis and photosynthetic gene expression^{25–27}. Here we find a role for brassinosteroid in stomatal production, which must be coordinated with other developmental processes to optimize photosynthetic and water-use efficiency. Notably, brassinosteroid represses light-responsive gene expression and chloroplast development mainly through the BZR1-mediated transcriptional network^{26,27}, but represses stomatal development through a BZR1-independent GSK3–MAPK crosstalk mechanism. Both GSK3 and MAPK are highly conserved in all eukaryotes, but it remains to be seen whether GSK3 directly inactivates MAPKKK

proteins in animals. This GSK3–MAPK connection has the potential to act in multiple receptor kinase-mediated signalling pathways, mediating crosstalk between these pathways in plants. The stronger stomata-clustering phenotype of *bsu-q* and suppression of *er1 erl2* stomata phenotypes by bikinin raise a possibility that members of the BSU1 and GSK3 families mediate signalling by the ERECTA family receptor kinases. However, the signals from BRI1 and ERECTA family must be partitioned differently downstream so that BRI1 controls GSK3 regulation of both BZR1 and YDA but ERECTA family mainly controls the GSK3 inactivation of YDA (Fig. 3k), because *er1 erl2* had no obvious effect on brassinosteroid-regulated BZR1 phosphorylation (Supplementary Fig. 13). Similar mechanisms and components might also be used by additional signalling pathways, such as the innate immunity pathway downstream of the FLS2 receptor kinase, which shares the BAK1 co-receptor²⁸ and downstream components MPK3 and MPK6 with BRI1 (ref. 23). In support of such an idea, overexpression of a GSK3-like kinase reduced the pathogen-induced activation of MPK3 and MPK6 (ref. 29). How signalling specificity is maintained when multiple pathways share the same components is a question for future study, and studies of the brassinosteroid model system will probably shed light on the hundreds of plant receptor kinases and their crosstalk during plant responses to complex endogenous and environmental cues.

METHODS SUMMARY

Stomatal quantification. Cotyledons of 8-day-old seedlings were cleared in ethanol with acetic acid (ratio of 19:1, v/v) and mounted on slides in Hoyer's solution (see ref. 22). Two to four images at $\times 400$ magnification ($180\ \mu\text{m}^2$) were captured per cotyledon from central regions of abaxial leaves. Guard cells, meristemoids, GMCs and pavement cells were counted. Statistical analysis was performed by Sigmaplot software (Systat Software). For treatment with bikinin²⁰, seedlings were grown on half-strength Murashige and Skoog (MS) medium containing dimethylsulphoxide (DMSO) or 30 μM bikinin (+10 μM oestradiol for HOPAI1-inducible lines) for 8 days before stomata were analysed.

Biochemical assays. To test the bikinin effect on YDA–Myc phosphorylation, homozygous *YDA-4Myc* plants were grown on 1/2 MS medium containing 2 μM BRZ (BRASSINAZOLE, an inhibitor of BR synthesis) for 5 days and treated with 30 μM bikinin or 2 μM BRZ solution for 30 min with gentle agitation. Yeast two-hybrid, *in vitro* interaction and kinase assays^{9,12}, and in-gel kinase assays³⁰ were carried out as described previously. Details of methods are available in the Supplementary Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 18 March; accepted 13 December 2011.

Published online 5 February 2012.

1. Bergmann, D. C. & Sack, F. D. Stomatal development. *Annu. Rev. Plant Biol.* **58**, 163–181 (2007).
2. Dong, J. & Bergmann, D. C. Stomatal patterning and development. *Curr. Top. Dev. Biol.* **91**, 267–297 (2010).
3. Thummel, C. S. & Chory, J. Steroid signaling in plants and insects—common themes, different pathways. *Genes Dev.* **16**, 3113–3129 (2002).
4. Kim, T. W. & Wang, Z. Y. Brassinosteroid signal transduction from receptor kinases to transcription factors. *Annu. Rev. Plant Biol.* **61**, 681–704 (2010).
5. Kinoshita, T. *et al.* Binding of brassinosteroids to the extracellular domain of plant receptor kinase BRI1. *Nature* **433**, 167–171 (2005).
6. Hothorn, M. *et al.* Structural basis of steroid hormone perception by the receptor kinase BRI1. *Nature* **474**, 467–471 (2011).
7. She, J. *et al.* Structural insight into brassinosteroid perception by BRI1. *Nature* **474**, 472–476 (2011).
8. Tang, W. *et al.* BSKs mediate signal transduction from the receptor kinase BRI1 in *Arabidopsis*. *Science* **321**, 557–560 (2008).
9. Kim, T.-W. *et al.* Brassinosteroid signal transduction from cell-surface receptor kinases to nuclear transcription factors. *Nature Cell Biol.* **11**, 1254–1260 (2009).

10. Tang, W. *et al.* PP2A activates brassinosteroid-responsive gene expression and plant growth by dephosphorylating BZR1. *Nature Cell Biol.* **13**, 124–131 (2011).
11. Li, J. & Nam, K. H. Regulation of brassinosteroid signaling by a GSK3/SHAGGY-like kinase. *Science* **295**, 1299–1301 (2002).
12. He, J. X., Gendron, J. M., Yang, Y., Li, J. & Wang, Z. Y. The GSK3-like kinase BIN2 phosphorylates and destabilizes BZR1, a positive regulator of the brassinosteroid signaling pathway in *Arabidopsis*. *Proc. Natl Acad. Sci. USA* **99**, 10185–10190 (2002).
13. Saibo, N. J., Vriezen, W. H., Beemster, G. T. & Van Der Straeten, D. Growth and stomata development of *Arabidopsis* hypocotyls are controlled by gibberellins and modulated by ethylene and auxins. *Plant J.* **33**, 989–1000 (2003).
14. Kang, C. Y., Lian, H. L., Wang, F. F., Huang, J. R. & Yang, H. Q. Cryptochromes, phytochromes, and COP1 regulate light-controlled stomatal development in *Arabidopsis*. *Plant Cell* **21**, 2624–2641 (2009).
15. Wang, H., Ngwenyama, N., Liu, Y., Walker, J. C. & Zhang, S. Stomatal development and patterning are regulated by environmentally responsive mitogen-activated protein kinases in *Arabidopsis*. *Plant Cell* **19**, 63–73 (2007).
16. Lampard, G. R., Macalister, C. A. & Bergmann, D. C. *Arabidopsis* stomatal initiation is controlled by MAPK-mediated regulation of the bHLH SPEECHLESS. *Science* **322**, 1113–1116 (2008).
17. Rowe, M. H. & Bergmann, D. C. Complex signals for simple cells: the expanding ranks of signals and receptors guiding stomatal development. *Curr. Opin. Plant Biol.* **13**, 548–555 (2010).
18. Szekeres, M. *et al.* Brassinosteroids rescue the deficiency of CYP90, a cytochrome P450, controlling cell elongation and de-etiolation in *Arabidopsis*. *Cell* **85**, 171–182 (1996).
19. Wang, Z. Y. *et al.* Nuclear-localized BZR1 mediates brassinosteroid-induced growth and feedback suppression of brassinosteroid biosynthesis. *Dev. Cell* **2**, 505–513 (2002).
20. De Rybel, B. *et al.* Chemical inhibition of a subset of *Arabidopsis thaliana* GSK3-like kinases activates brassinosteroid signaling. *Chem. Biol.* **16**, 594–604 (2009).
21. Rozhon, W., Mayerhofer, J., Petutschnig, E., Fujioka, S. & Jonak, C. ASK0, a group-III *Arabidopsis* GSK3, functions in the brassinosteroid signalling pathway. *Plant J.* **62**, 215–223 (2010).
22. Bergmann, D. C., Lukowitz, W. & Somerville, C. R. Stomatal development and pattern controlled by a MAPKK kinase. *Science* **304**, 1494–1497 (2004).
23. Zhang, J. *et al.* A *Pseudomonas syringae* effector inactivates MAPKs to suppress PAMP-induced immunity in plants. *Cell Host Microbe* **1**, 175–185 (2007).
24. Kanaoka, M. M. *et al.* SCREAM/ICE1 and SCREAM2 specify three cell-state transitional steps leading to *Arabidopsis* stomatal differentiation. *Plant Cell* **20**, 1775–1785 (2008).
25. Li, J., Nagpal, P., Vitart, V., McMorris, T. C. & Chory, J. A role for brassinosteroids in light-dependent development of *Arabidopsis*. *Science* **272**, 398–401 (1996).
26. Sun, Y. *et al.* Integration of brassinosteroid signal transduction with the transcription network for plant growth regulation in *Arabidopsis*. *Dev. Cell* **19**, 765–777 (2010).
27. Luo, X.-M. *et al.* Integration of light and brassinosteroid signaling pathways by a GATA transcription factor in *Arabidopsis*. *Dev. Cell* **19**, 872–883 (2010).
28. Chinchilla, D. *et al.* A flagellin-induced complex of the receptor FLS2 and BAK1 initiates plant defence. *Nature* **448**, 497–500 (2007).
29. Wrzaczek, M., Rozhon, W. & Jonak, C. A proteasome-regulated glycogen synthase kinase-3 modulates disease response in plants. *J. Biol. Chem.* **282**, 5249–5255 (2007).
30. Zhang, S. & Klessig, D. F. Salicylic acid activates a 48-kD MAP kinase in tobacco. *Plant Cell* **9**, 809–824 (1997).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank K. Torii for providing seeds of the *er1 erl2* triple mutant and *scrm-D* mutant, J.-M. Zhou for providing the *HOPAI1* transgenic line and J. Li for providing the *bin2-3 bil1 bil2* triple mutant. Research was primarily supported by a grant from the National Institutes of Health (R01GM066258), and partially supported by the US Department of Energy (DE-FG02-08ER15973) and the Herman Frasch Foundation. D.C.B. is an investigator of the Howard Hughes Medical Institute.

Author Contributions T.-W.K. performed all experiments. T.-W.K. and Z.-Y.W. designed the experiments, analysed data and wrote the manuscript. M.M. cloned complementary DNAs of BSL2 and BSL3, and BSL2pro-BSL2. D.C.B. contributed materials and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to Z.-Y.W. (zywang24@stanford.edu).

METHODS

Materials and growth conditions. All mutants are in the Columbia ecotype except *yda* Y295 (C24 ecotype)²², CA-YDA (Ler ecotype)²² and *bin2-3 bil1 bil2* triple mutant obtained from J. Li (Ws ecotype)³¹. The *erecta* triple mutant *er105 erl1-2 erl2-1* (ref. 32) and *scrm-D* (ref. 24) were obtained from K. Torii. J.-M. Zhou provided seeds of oestradiol-inducible *HOPAI1* transgenic plants²³. For all analyses, *Arabidopsis* seedlings were grown on MS agar medium for 8 days under continuous light in Percival growth chamber at 22 °C.

Stomatal quantification. Cotyledons of 8-day-old seedlings were cleared in ethanol with acetic acid and mounted on slides in Hoyer's solution (see ref. 22). Two to four images at $\times 400$ magnification ($180\ \mu\text{m}^2$) were captured per cotyledon from central regions of abaxial leaves. Guard cells, meristemoids, GMCs and pavement cells were counted. Statistical analysis was performed by SigmaPlot software (Systat Software). For treatment with bikinin²⁰, seedlings were grown on half-strength MS medium containing DMSO or 30 μM bikinin (+10 μM estradiol for *HOPAI1*-inducible lines) for 8 days before stomata were analysed.

Plasmids. For cloning MBP-185/322, a partial cDNA was amplified from a YDA cDNA clone using primers (forward; 5'-caccAGTAACAAAACTCAGCTGAGATGTTT-3', reverse; 5'-AGAGCTAG GACCAGGGCTTGTCATTCT-3'), cloned into pENTR-SD-D-TOPO vector (Invitrogen) and then subcloned into the gateway-compatible pMALc2 vector (New England Biolab). For expression in plants, cDNA entry clones of YDA and CA-YDA were subcloned into a gateway-compatible 35S::4myc-6His vector constructed in the pCambia 1390 vector. *BSL2* cDNA in the pENTR vector was subcloned into Gateway-compatible pEarley-101 vector³³ to generate 35S::*BSL2-YFP*.

Overlay assay. To test the interaction of YDA and BIN2 *in vitro*, a gel blot separating MBP, MBP-CDG1 (a protein kinase used as a negative control) and MBP-YDA was incubated with 20 μg GST-BIN2 in 5% non-fat dry milk/PBS buffer and washed four times. The blot was then probed with HRP-conjugated anti-GST antibody (Santa Cruz Biotechnology).

In vitro kinase assay. Induction and purification of proteins expressed from *Escherichia coli* was performed as described previously¹². For Fig. 3d, e, 1 μg of GST-BIN2 or 0.5 μg of MBP-BIN2 was incubated with 1 μg of MBP-YDA or MBP-185/322 in the kinase buffer (20 mM Tris, pH 7.5, 1 mM MgCl_2 , 100 mM NaCl and 1 mM DTT) containing 100 μM ATP and 10 μCi [^{32}P] γ -ATP at 30 °C for 3 h. To examine whether BIN2 inhibits YDA activity, equal amounts of MBP-YDA were pre-incubated with GST-BIN2 or GST-mBIN2 (M115A) for 2 h. Pre-incubated MBP-YDA was subsequently purified using glutathione beads and amylose beads to remove GST-BIN2 or GST-mBIN2. Purified YDA was then incubated with GST-mMKK4 (K108R), 10 μCi [^{32}P] γ -ATP and 10 μM bikinin (to inhibit any residual BIN2) at 30 °C for 3 h. YDA kinase activity towards mMKK4 was analysed by SDS-PAGE followed by autoradiography.

In-gel kinase assay. The in-gel kinase assay was performed as described previously³⁰, with some modifications. Total proteins were extracted with buffer containing 50 mM Tris, pH 7.5, 150 mM NaCl, 5% Glycerol, 1% Triton X-100, 1 mM

phenylmethylsulphonyl fluoride (PMSF), 1 μM E-64, 1 μM bestatin, 1 μM pepstatin and 2 μM leupeptin. Supernatant obtained from 12,000 r.p.m. centrifugation was quantified by Bradford protein assay. Equal amounts of protein (40 μg) were loaded on 10% SDS-PAGE gel embedded with 0.2 mg ml^{-1} of myelin basic protein. After electrophoresis, SDS was removed by incubation with washing buffer (25 mM Tris, pH 7.5, 0.5 mM DTT, 5 mM NaF, 0.1 mM Na_3VO_4 , 0.5 mg ml^{-1} bovine serum albumin and 0.1% Triton X-100) with three buffer exchanges at 22 °C for 1.5 h. The gel was incubated with renaturation buffer (25 mM Tris, pH 7.5, 0.5 mM DTT, 5 mM NaF and 0.1 mM Na_3VO_4) at 4 °C overnight with four buffer exchanges. After pre-incubation with 100 ml of kinase reaction buffer without ATP for 30 min, the gel was incubated with 30 ml of kinase reaction buffer (25 mM Tris, pH 7.5, 2 mM EGTA, 12 mM MgCl_2 , 1 mM DTT, 0.1 mM Na_3VO_4 , 200 nM ATP and 50 μCi [^{32}P] γ -ATP) for 1.5 h. The gel was washed with solution containing 5% trichloroacetic acid (w/v) and 1% potassium pyrophosphate (w/v) four times for 2–3 h. Dried gel was exposed with phosphor screen followed by phospho-imager analysis.

Transient interaction assays and analysis of bikinin effects on YDA in transgenic plants. *Agrobacterium* GV3101 strains transformed with 35S::CA-YDA-4Myc-6His or 35S::YDA-4Myc-6His constructs were alone or co-infiltrated with 35S::BIN2-YFP expressing *Agrobacterium* into *N. benthamiana* leaves as described previously⁹. After 36 h, protein extracts were prepared from *N. benthamiana* leaves in immunoprecipitation buffer containing 50 mM Tris, pH 7.5, 150 mM NaCl, 5% Glycerol, 1% Triton X-100, 1 mM PMSF, 1 μM E-64, 1 μM bestatin, 1 μM pepstatin and 2 μM leupeptin. Supernatant obtained from 20,000g centrifugation was incubated with anti-YFP-antibody-bound protein A beads for 1 h. Beads were washed 5 times with immunoprecipitation buffer containing 0.2% Triton X-100. Immunoprecipitated proteins were eluted with 2 \times SDS Laemmli buffer, separated on SDS-PAGE and subjected to immunoblotting using anti-Myc antibody (Abcam) and anti-YFP antibody.

For transgenic *Arabidopsis* plants, wild-type *Arabidopsis* was transformed with *Agrobacterium* containing 35S::YDA-4Myc-6His or 35S::*BSL2-YFP* construct by floral dip. Hygromycin or Basta-resistant T1 plants were screened by immunoblot using anti-Myc or anti-YFP antibody, respectively.

To test the bikinin effect on YDA-Myc phosphorylation, homozygous YDA-4Myc plants were grown on half-strength MS medium containing 2 μM BRZ for 5 days and treated with 30 μM bikinin or 2 μM BRZ solution for 30 min with gentle agitation. YDA-Myc was analysed by immunoblot.

31. Yan, Z., Zhao, J., Peng, P., Chihara, R. K. & Li, J. BIN2 functions redundantly with other *Arabidopsis* GSK3-like kinases to regulate brassinosteroid signaling. *Plant Physiol.* **150**, 710–721 (2009).
32. Shpak, E. D., McAbee, J. M., Pillitteri, L. J. & Torii, K. U. Stomatal patterning and differentiation by synergistic interactions of receptor kinases. *Science* **309**, 290–293 (2005).
33. Earley, K. W. *et al.* Gateway-compatible vectors for plant functional genomics and proteomics. *Plant J.* **45**, 616–629 (2006).

Single-molecule imaging of DNA pairing by RecA reveals a three-dimensional homology search

Anthony L. Forget^{1,2} & Stephen C. Kowalczykowski^{1,2}

DNA breaks can be repaired with high fidelity by homologous recombination. A ubiquitous protein that is essential for this DNA template-directed repair is RecA¹. After resection of broken DNA to produce single-stranded DNA (ssDNA), RecA assembles on this ssDNA into a filament with the unique capacity to search and find DNA sequences in double-stranded DNA (dsDNA) that are homologous to the ssDNA. This homology search is vital to recombinational DNA repair, and results in homologous pairing and exchange of DNA strands. Homologous pairing involves DNA sequence-specific target location by the RecA-ssDNA complex. Despite decades of study, the mechanism of this enigmatic search process remains unknown. RecA is a DNA-dependent ATPase, but ATP hydrolysis is not required for DNA pairing and strand exchange^{2,3}, eliminating active search processes. Using dual optical trapping to manipulate DNA, and single-molecule fluorescence microscopy to image DNA pairing, we demonstrate that both the three-dimensional conformational state of the dsDNA target and the length of the homologous RecA-ssDNA filament have important roles in the homology search. We discovered that as the end-to-end distance of the target dsDNA molecule is increased, constraining the available three-dimensional (3D) conformations of the molecule, the rate of homologous pairing decreases. Conversely, when the length of the ssDNA in the nucleoprotein filament is increased, homology is found faster. We propose a model for the DNA homology search process termed 'intersegmental contact sampling', in which the intrinsic multivalent nature of the RecA nucleoprotein filament is used to search DNA sequence space within 3D domains of DNA, exploiting multiple weak contacts to rapidly search for homology. Our findings highlight the importance of the 3D conformational dynamics of DNA, reveal a previously unknown facet of the homology search, and provide insight into the mechanism of DNA target location by this member of a universal family of proteins.

The mechanism by which the RecA family of DNA strand exchange proteins (which include T4 UvsX, archaeal RadA and eukaryotic Rad51) locate DNA sequence identity is unknown. Ensemble studies have constrained possible mechanisms by establishing that ATP hydrolysis is not needed^{3,4} and 1D sliding is not operative⁵. Consequently, the manner by which the RecA nucleoprotein filament promotes the efficient, rapid and accurate search for homology has remained undefined for decades⁶. Single-molecule methods have the potential to provide new insight into this long-standing question. In fact, magnetic tweezer experiments showed that the endpoint of homologous pairing can be detected as a change in the length of a single dsDNA target molecule^{7,8}. However, the mechanism by which homology was found and DNA pairing occurred was not shown. Therefore, we sought to directly observe the manner by which RecA nucleoprotein filaments locate their homologous target in dsDNA.

Initially we attempted to directly observe fluorescent RecA nucleoprotein filaments interacting with bacteriophage λ dsDNA in real time by using total internal reflected fluorescence microscopy (TIRFM)⁹. Fully homologous fluorescent ssDNA that was complementary to

three different loci of λ DNA (Fig. 1A) was generated by incorporation of 5-(3-aminoallyl) dUTP into ssDNA using polymerase chain reaction (PCR), followed with covalent attachment of ATTO565 (Supplementary Methods). RecA nucleoprotein filaments were assembled on these fluorescent ssDNA substrates in ensemble reactions containing ssDNA-binding protein (SSB) and the non-hydrolysable ATP analogue, ATP γ S (5'-O-3'-thiotriphosphate)⁴. ATP γ S was used to maintain the filament in its active form, eliminate filament disassembly and prevent dissociation of DNA pairing products^{7,10–12}. Using

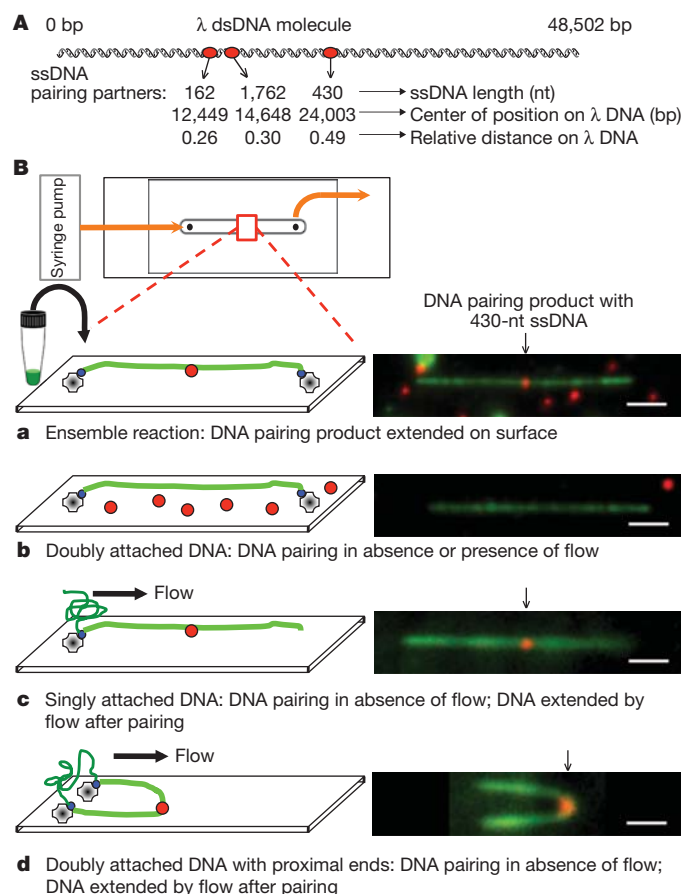


Figure 1 | DNA pairing by RecA, imaged using single-molecule TIRFM, indicates that the three-dimensional conformation of target dsDNA is important in the homology search. **A**, DNA substrates. **B**, DNA pairing between λ DNA (green) and RecA filament assembled on 430-nucleotide (nt) ssDNA (red). The ensemble reaction was examined by TIRFM (**B, a**). In *in situ* reactions dsDNA was attached before pairing; doubly attached extended DNA (**B, b**), singly attached DNA (**B, c**) and doubly attached DNA with ends in proximity (**B, d**). Homologously paired products were observed in **B, c** and **B, d** when DNA was relaxed by stopping flow and then extended by flow for visualization. Scale bars, 2.4 μ m.

¹Department of Microbiology, University of California, Davis, California 95616-8665, USA. ²Department of Molecular and Cellular Biology, University of California, Davis, California 95616-8665, USA.

biochemical assays, we confirmed that the fluorescent ssDNA that was generated by this procedure was functional for RecA-mediated DNA pairing (Supplementary Fig. 1). The λ dsDNA, biotinylated at each end, was attached under flow to the interior surface of a single-channel microfluidic device (flowcell) (Fig. 1B). Owing to sequential attachment of each end to the streptavidin-coated surface, most DNA molecules were extended to nearly ($\sim 80\%$) B-form length, and extension could be maintained in the absence of flow (Fig. 1B, a, b).

To confirm DNA pairing at the homologous λ DNA target site, reactions were conducted under ensemble conditions, and products were extended on the surface of a flowcell for analysis by single-molecule, two-colour TIRFM; dsDNA was imaged by YOYO1 binding (green) and ssDNA by ATTO565 (red). DNA pairing products were observed; the sites of interaction coincided with the region of homology within the λ DNA molecule (Fig. 1B, a). For the 430-nucleotide ssDNA, all bound fluorescent ssDNA RecA filaments were at the homologous locus (observed fractional distance 0.51 ± 0.02 ; $n = 21$; Supplementary Fig. 2).

Next, we attempted to detect homologous pairing in real time using single-molecule TIRFM. Preformed RecA nucleoprotein filaments were introduced into a flowcell to which λ DNA molecules were tethered, buffer flow was stopped, and the reaction was monitored in real time (Fig. 1B, b). Although the dsDNA was readily visible, we failed to observe any interaction between the fluorescent nucleoprotein filaments and extended λ DNA, even for reaction periods longer than 1 h. However, we noticed that in addition to the desired doubly tethered extended λ DNA molecules, some DNA molecules were attached only by one end (Fig. 1B, c). When flow was stopped to score pairing with the doubly tethered λ DNA molecules, these singly tethered molecules relaxed to a randomly coiled state. Unexpectedly, when these unconstrained DNA molecules were subsequently re-extended by buffer flow, 80% ($n = 20$) revealed a stable pairing product (Fig. 1B, c). This finding suggested that either a free DNA end or a random coiled DNA was needed for pairing. In the same field of view, there were also λ DNA molecules that had both ends attached, but at a relatively close end-to-end distance (Fig. 1B, d). When the flow was stopped, we observed that these molecules also participated in homologous pairing during the time that flow was off, demonstrating that a free DNA end was not required. These unanticipated results revealed that DNA pairing did not occur on DNA that was extended to near its entropic elastic limit, and suggested that the DNA homology search required the 3D states that are accessible in randomly coiled DNA. Collectively, they suggested that a coiled conformation of the target dsDNA is crucial.

To address this possibility, we developed an alternative single-molecule imaging strategy that permitted reproducible measurement of the effects of dsDNA conformational structure, unperturbed by flow, on the DNA homology search process. This method uses a specialized flowcell (Fig. 2A), two optical laser traps operated in position-clamp mode, epifluorescent detection, fluorescent RecA-ssDNA filaments and a λ DNA dumbbell (a single λ DNA molecule with a 1- μ m polystyrene bead attached at each end¹³ (Supplementary Methods)). The DNA pairing assay was performed *in situ* using the dsDNA dumbbell target, and the dual optical trap configuration was used to reliably vary the end-to-end distance of the dsDNA. The flowcell has four channels and a flow-free reservoir. Movement of DNA dumbbells between channels of the flowcell was accomplished through stage translation, and manipulation of optical traps relative to one another was accomplished using a steering mirror controlling one of the traps. Each experiment (Fig. 2B and Supplementary Movie 1) consisted of six steps: first, in channel one, a streptavidin-coated bead was trapped in each of the two optical traps (Fig. 2B, a); second, the beads were moved to channel two to capture a λ dsDNA molecule (biotinylated on both ends and stained with YOYO1) on one bead (Fig. 2B, b); third, the beads were moved into channel three, and by independent steering of a trap, the distal end of the DNA was attached

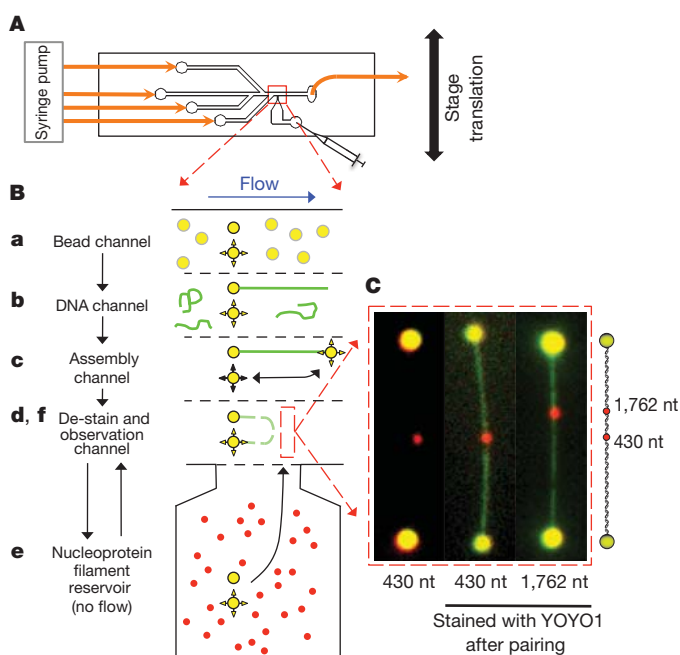


Figure 2 | Visualization of RecA-promoted DNA pairing with an individual optically trapped DNA dumbbell, imaged by epifluorescence. A, Four-channel flowcell with a flow-free reservoir. B, DNA dumbbell assembly and RecA-pairing reaction: two beads (yellow) are trapped (B, a); a λ DNA molecule (green) is captured on one bead (B, b); the free DNA end is captured with the second bead using a steerable optical trap (B, c); the centre-to-centre bead distance is set and YOYO1 is removed (B, d, de-stain); the DNA dumbbell is incubated in reservoir with RecA nucleoprotein filaments (red) (B, e) and DNA is extended to visualize products (B, f). C, Images of pairing products with 430- and 1,762-nucleotide nucleoprotein filaments.

to the second bead (Fig. 2B, c); fourth, the DNA-dumbbell was moved to the dye-free channel for de-staining, and the end-to-end distance was fixed (Fig. 2B, d); fifth, the DNA-dumbbell was moved to the flow-free reservoir containing the fluorescent ssDNA-RecA filaments (Fig. 2B, e); and sixth, after a defined incubation time, the DNA dumbbell was moved back to channel four, which is free of nucleoprotein filaments, extended to its contour length ($\sim 16 \mu$ m) and examined for DNA pairing products (Fig. 2B, f).

Shown in Fig. 2C are representative products of reactions in which the DNA dumbbells were initially held at a centre-to-centre bead distance of 2μ m and incubated for 2 min in the reservoir that contained RecA nucleoprotein filaments. For the two homologous ssDNA nucleoprotein filaments shown (430 nucleotides and 1,762 nucleotides), the pairing is clearly at the homologous locus. For a 2 min incubation with dsDNA at a bead-to-bead distance of 2μ m and the 430-nucleotide substrate, 90% of the dsDNA molecules ($n = 29$) contained a nucleoprotein filament stably bound to the expected region of homology (Fig. 3a). To determine the effect of end-to-end distance (that is, 3D conformation) on the RecA-mediated DNA pairing reaction, the reactions were performed at increasing bead separations (Fig. 3a). As the bead distance was increased from 2μ m to 8μ m, the efficiency of DNA pairing decreased to near zero, extrapolating to zero at $\sim 9 \mu$ m; for comparison, in the TIRFM experiments in which no DNA pairing was detected *in situ*, the DNA end-to-end distance was $\sim 13 \mu$ m.

We compared the time course of homologous pairing for fixed centre-to-centre bead distances of 2μ m and 6μ m (Fig. 3b) to determine the effect of decreasing DNA conformational states on the rate of the reaction. For the 2μ m separation, the rate of DNA pairing increased with a half-time of ~ 30 s and approached a yield of 100%. When the separation was increased to 6μ m, the rate slowed fourfold to a half-time of ~ 125 s, but nevertheless approached a yield of 100%

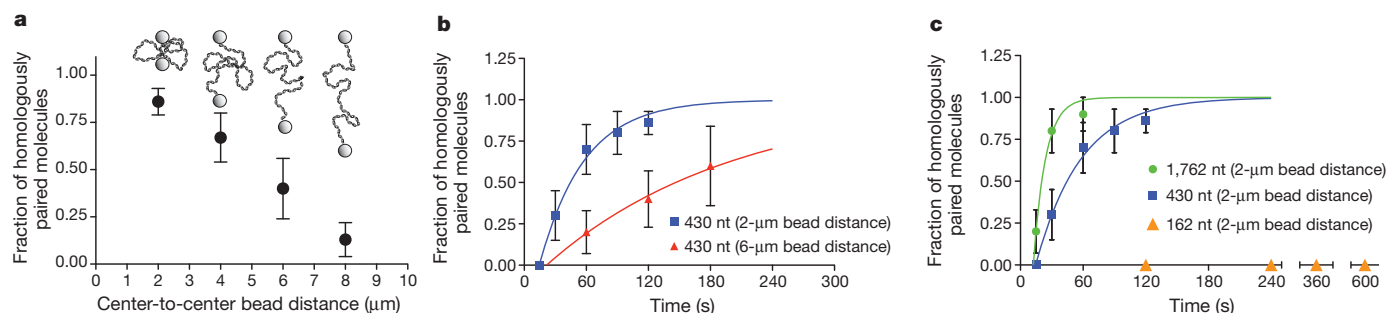


Figure 3 | DNA three-dimensional conformation and nucleoprotein filament length contribute to the homology search. **a**, Effect of DNA end-to-end distance; 430-nucleotide substrate (2 min). Error bars, s.e.m. from multiple experiments ($n = 10$ to 29). **b**, Time course; 430-nucleotide substrate: 2- μm (squares) and 6- μm (triangles) bead separation; respective pairing rates, 0.023

(± 0.002) s^{-1} ($n \geq 10$) and 0.0056 (± 0.0006) s^{-1} ($n \geq 5$). **c**, Effect of ssDNA length; 162 nucleotides (triangles; $n = 5, 6, 4$ and 2 at the times indicated), 430 nucleotides (squares; same data as Fig. 3b; $n \geq 10$) and 1,762 nucleotides (circles; $n \geq 10$); error bars, s.e.m.; 2- μm separation; respective rates: zero, 0.023 (± 0.002) s^{-1} , and 0.086 (± 0.026) s^{-1} .

(Fig. 3b). To establish the kinetic reaction order, we conducted single-molecule DNA pairing assays as a function of RecA nucleoprotein filament concentration (Supplementary Fig. 3). The reaction rate was independent of nucleoprotein filament concentration, showing that DNA pairing under these conditions is not diffusion limited, but that it is limited instead by a rate-determining unimolecular step as in the ensemble studies¹⁴. However, the pairing rate was dependent on dsDNA conformation and therefore was not dependent on the sequence recognition step itself.

To understand the nature of the complex that limits the rate of DNA pairing, we varied the length of RecA nucleoprotein filaments. Shown in Fig. 3c is a comparison of the time courses for 162-, 430- and 1,762-nucleotide nucleoprotein filaments. Increasing the ssDNA length approximately fourfold, from 430 to 1,762 nucleotides, increased the observed rate of pairing approximately 3.8-fold. However, when the length of the ssDNA was decreased to 162 nucleotides, we did not observe any stably bound homology paired products after incubations for 10 min at the closest bead-to-bead distance possible (2 μm), despite this substrate being active in ensemble DNA pairing reactions (Supplementary Fig. 2). We conclude that the length of the RecA nucleoprotein filament is a crucial factor in the rate-limiting step of homologous pairing.

In addition to the anticipated stable, homology paired end products, short-lived non-homologous interactions were observed (Fig. 4a). These events, which occurred outside of the homologous regions, were relatively unstable and dissociated during the movement of the molecule from the reservoir to the observation channel, during the separation of beads or after the λ DNA molecule was extended (Supplementary Movie 2). These heterologous events lasted no more than a few tens of seconds and never persisted on a timescale of minutes. When the molecules from the 2- μm data set were analysed, 22% of the reactions with the 430-nucleotide ssDNA and 40% of reactions with the 1,762-nucleotide ssDNA had these unstable heterologously paired intermediates (Fig. 4b), and for the 162-nucleotide ssDNA, only 1 heterologously bound filament was seen out of 28 molecules.

Some intermediates of the pairing process had a second filament bound non-specifically to spatially separated regions of the λ DNA molecule. For such a heterologously bound nucleoprotein filament, when the relaxed DNA molecule was moved into the observation channel and the beads were separated for observation, the existence of a loop could be inferred from a sudden recoil of the homology paired spot. As the beads were separated, the weaker of the two heterologous interactions was released, and there was a simultaneous movement ('jump') of the fluorescence at the homologous pairing locus (Fig. 4a and Supplementary Movie 3) resulting from the release of DNA that was constrained in the loop. Approximately 12% ($n = 50$) of the DNA dumbbells showed loop release events for the 430-nucleotide nucleoprotein filament and, consistent with expectations,

when the length of the nucleoprotein filament was increased to 1,762 nucleotides, the number of molecules with transient loop structures increased to 47% ($n = 30$) (Fig. 4c).

Our results clearly establish that both the 3D conformation of dsDNA and the length of the nucleoprotein filament are important determinants of the rate for DNA homologous pairing. These findings lead us to propose a model termed 'intersegmental contact sampling' to describe the search for homology by a RecA nucleoprotein filament (Fig. 4d). One of the key features of the model is that the RecA nucleoprotein filament has a polyvalent interaction surface that is capable of binding simultaneously and non-specifically, but weakly, with non-contiguous segments of dsDNA. The second related feature of this

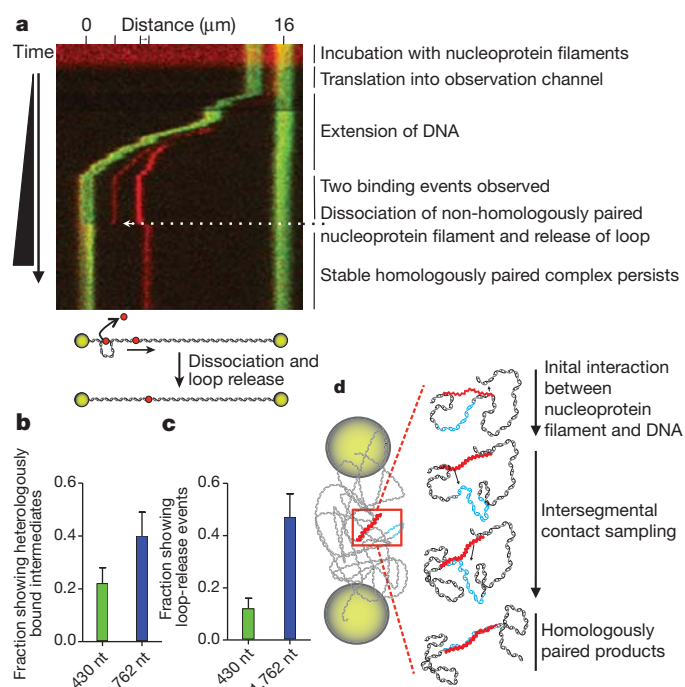


Figure 4 | RecA nucleoprotein filaments exhibit transient non-homologous interactions and loop-release events. **a**, Kymograph of DNA dumbbell during bead separation (Fig. 2B, f). Distance scale (top) and tick marks show positions of beads (green) and nucleoprotein filaments (red); illustration depicts dissociation of heterologously bound filament. **b**, **c**, Fraction of dsDNA dumbbells with non-homologously bound intermediates (**b**) and loop release events (**c**); 430-nucleotide (blue) and 1,762-nucleotide (green) filaments; $n = 50$ and 30, respectively. Error bars, s.e.m. **d**, Model for RecA homology search by intersegmental contact sampling; for simplicity, only two simultaneous points of interaction are depicted.

model is that 3D conformational entropy of the dsDNA greatly enhances the probability that DNA sequence homology will be found through iterated homology sampling, using multiple weak contacts, by this polyvalent filament. This model is compatible both with our key experimental findings, which we expect would apply to the search in the presence of ATP as well, and with the involvement of heterologously bound intermediates that have been inferred from biochemical studies^{15,16}. Our data show that dsDNA extended to near contour length fails to produce homologously paired products. This observation provides an explanation for the observation that the formation of stable DNA pairing products in single-molecule studies using magnetic tweezers required negative plectonemic supercoils in the DNA target^{7,8}. By contrast, when a ssDNA–RecA filament was extended to near its contour length, homologous pairing with fully homologous coiled dsDNA occurred⁷, which is compatible with our finding that the coiled structure of dsDNA is essential to the homology search. Here we established that as the end-to-end distance of the dsDNA was decreased, allowing it to assume a more random coil-like 3D conformation, the rate of DNA pairing increased because the local DNA concentration increases, and the likelihood that DNA segments will be in close proximity also greatly increases. The increased local DNA concentration results in a greater statistical probability that a single nucleoprotein filament can simultaneously interact with and sample multiple regions of the same DNA molecule. This, in turn, is manifest as a kinetically more efficient homology sampling process. In further support of the intersegmental contact sampling model, when the length of the ssDNA in the nucleoprotein filament is increased, the observed rate of pairing, as well as the number of nucleoprotein filaments with multiple, transient, heterologous intersegmental interactions is increased. This shows that longer nucleoprotein filaments can simultaneously and independently sample more segments of the target dsDNA than shorter nucleoprotein filaments. Kinetically, our findings are consistent with the following two-step scheme:



where K_{het} is the equilibrium constant for the binding of a RecA nucleoprotein filament (NPF) to heterologous dsDNA (the kinetic steps comprising K_{het} are rapid compared to k_s) and k_s is the rate-limiting unimolecular rate constant for intersegmental homology searching step within the dsDNA molecule or domain. In general, this kinetic formalism predicts a hyperbolic dependence of homologous pairing on the component concentrations unless the equilibrium constant for formation of the heterologous complex is large; when this is case, the observed rate is defined by the first-order rate constant, k_s . Given that the rate of target location is independent of nucleoprotein filament concentration, this implies that the heterologously bound complex is saturated at a filament concentration of 100 pM (Supplementary Fig. 3), placing a limit on the apparent equilibrium dissociation constant of <10 pM (that is, $K_{\text{het}} > 10^{11} \text{ M}^{-1}$). In the context of this kinetic model, values for k_s are defined by the experiments in Fig. 3b, c, which show that the rate of the intersegmental homology search decreases fourfold when the DNA end-to-end distance increases from 1 μm to 5 μm and increases approximately fourfold when the ssDNA length increases approximately fourfold. The correlation of rate with the length of ssDNA suggests that the intradomain search is enhanced proportionately by the increase in either heterologous contacts or the reach of the longer ssDNA. In many regards, the homology search by RecA has parallels to target location by sequence-specific DNA-binding proteins, with the notable exception that the specificity of the RecA filament is determined by the sequence of the associated ssDNA. Seminal work on the DNA target selection by transcriptional regulatory proteins identified sliding, hopping and intersegmental transfer as potentially facilitating mechanisms^{17,18}. Here we have established intersegmental transfer as the operative pathway used by RecA to find DNA sequence homology;

this behaviour is distinct from the sliding and hopping used to enhance the rate of target location by most regulatory proteins, which are typically univalent or bivalent with regard to site binding¹⁸. Our approach now provides a framework for future studies on the previously mysterious homology search by recombination proteins. It is applicable to studies of more complex systems such as eukaryotic Rad51, as it can provide insight into the function of the many accessory proteins that enhance DNA pairing⁹. Finally, the imaging strategy and flow-free cell design can easily be adapted to visualize target location and mechanism of processes as diverse as DNA replication and repair, RNA interference, transcription and protein translation, in which the 3D conformations of nucleic acids are undoubtedly important.

METHODS SUMMARY

RecA and SSB were purified as described^{19,20}. Fluorescent ssDNA was prepared as detailed in the Supplementary Information. Nucleoprotein filaments were formed as described⁴ in SM buffer (25 mM Tris acetate (Tris-OAc) (pH 7.5), 1 mM DTT and 4 mM $\text{Mg}(\text{OAc})_2$), SSB (at a ratio of 1 SSB monomer to 11 nucleotides), 2 nM molecules fluorescent ssDNA, and 1 mM $\text{ATP}\gamma\text{S}$ were incubated for 10 min at 37 °C; RecA was added at 1 monomer per 1.7 nucleotides, and incubated 1 h. Nucleoprotein filaments were diluted to 0.2 nM before use.

For DNA pairing using TIRFM, biotinylated λ DNA (1 pM, molecules) in SM2 (SM with 50 mM NaCl) was bound to the flowcell and then washed to remove free DNA, and to attach the second DNA end. Reactions were started by addition of 0.2 nM nucleoprotein filaments. For ensemble experiments visualized by TIRFM, nucleoprotein filaments and λ DNA were incubated for 1 h (162-nucleotide substrate) or 30 min (430-nucleotide substrate) at 37 °C.

Visualization of RecA-mediated pairing with individual DNA dumbbells was performed at 37 °C. The flowcell was treated for 1 h with BSA (1 mg ml^{-1}) in SM3 (50 mM Tris-OAc (pH 8.2), 50 mM DTT, 1 mM $\text{Mg}(\text{OAc})_2$ and 15% sucrose). Biotinylated λ DNA and buffers were pumped into the flowcell at a linear flow rate of $\sim 100 \mu\text{m s}^{-1}$. Channels contained SM3, 18 fM streptavidin-coated polystyrene beads (1 μm , Bangs Laboratories) and 5 nM YOYO-1 (Invitrogen) (Fig. 2B, a); SM3, 100 nM YOYO1 and 10 pM (molecules) biotinylated λ DNA (Fig. 2B, b); SM3 (Fig. 2B, c); SM and 15% sucrose (Fig. 2B, d, f). The reaction reservoir contained 0.2 nM nucleoprotein filaments in SM with 15% sucrose and 0.5 mM $\text{ATP}\gamma\text{S}$ (Fig. 2B, e).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 12 September; accepted 12 December 2011.

Published online 8 February 2012.

- Kowalczykowski, S. C., Dixon, D. A., Eggleston, A. K., Lauder, S. D. & Rehauer, W. M. Biochemistry of homologous recombination in *Escherichia coli*. *Microbiol. Rev.* **58**, 401–465 (1994).
- Menetski, J. P. & Kowalczykowski, S. C. Interaction of recA protein with single-stranded DNA. Quantitative aspects of binding affinity modulation by nucleotide cofactors. *J. Mol. Biol.* **181**, 281–295 (1985).
- Kowalczykowski, S. C. & Krupp, R. A. DNA-strand exchange promoted by RecA protein in the absence of ATP: implications for the mechanism of energy transduction in protein-promoted nucleic acid transactions. *Proc. Natl Acad. Sci. USA* **92**, 3478–3482 (1995).
- Menetski, J. P., Bear, D. G. & Kowalczykowski, S. C. Stable DNA heteroduplex formation catalyzed by the *Escherichia coli* RecA protein in the absence of ATP hydrolysis. *Proc. Natl Acad. Sci. USA* **87**, 21–25 (1990).
- Adzuma, K. No sliding during homology search by RecA protein. *J. Biol. Chem.* **273**, 31565–31573 (1998).
- Kowalczykowski, S. C. Biochemistry of genetic recombination: energetics and mechanism of DNA strand exchange. *Annu. Rev. Biophys. Biophys. Chem.* **20**, 539–575 (1991).
- Fulconis, R., Miné, J., Bancaud, A., Dutreix, M. & Viovy, J. L. Mechanism of RecA-mediated homologous recombination revisited by single molecule nanomanipulation. *EMBO J.* **25**, 4293–4304 (2006).
- van der Heijden, T. et al. Homologous recombination in real time: DNA strand exchange by RecA. *Mol. Cell* **30**, 530–538 (2008).
- Forget, A. L. & Kowalczykowski, S. C. Single-molecule imaging brings Rad51 nucleoprotein filaments into focus. *Trends Cell Biol.* **20**, 269–276 (2010).
- McEntee, K., Weinstock, G. M. & Lehman, I. R. Binding of the recA protein of *Escherichia coli* to single- and double-stranded DNA. *J. Biol. Chem.* **256**, 8835–8844 (1981).
- Honigberg, S. M., Gonda, D. K., Flory, J. & Radding, C. M. The pairing activity of stable nucleoprotein filaments made from recA protein, single-stranded DNA, and adenosine 5'-(gamma-thio)triphosphate. *J. Biol. Chem.* **260**, 11845–11851 (1985).

12. Galletto, R., Amitani, I., Baskin, R. J. & Kowalczykowski, S. C. Direct observation of individual RecA filaments assembling on single DNA molecules. *Nature* **443**, 875–878 (2006).
13. van Mameren, J. *et al.* Counting RAD51 proteins disassembling from nucleoprotein filaments under tension. *Nature* **457**, 745–748 (2009).
14. Julin, D. A., Riddles, P. W. & Lehman, I. R. On the mechanism of pairing of single- and double-stranded DNA molecules by the recA and single-stranded DNA-binding proteins of *Escherichia coli*. *J. Biol. Chem.* **261**, 1025–1030 (1986).
15. Gonda, D. K. & Radding, C. M. By searching processively recA protein pairs DNA molecules that share a limited stretch of homology. *Cell* **34**, 647–654 (1983).
16. Tsang, S. S., Chow, S. A. & Radding, C. M. Networks of DNA and recA protein are intermediates in homologous pairing. *Biochemistry* **24**, 3226–3232 (1985).
17. Berg, O. G., Winter, R. B. & von Hippel, P. H. Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. *Biochemistry* **20**, 6929–6948 (1981).
18. Berg, O. G. in *The Biology of Nonspecific DNA Protein Interactions* (ed. Revzin, A.) 71–85 (CRC, 1990).
19. Mirshad, J. K. & Kowalczykowski, S. C. Biochemical characterization of a mutant RecA protein altered in DNA-binding loop 1. *Biochemistry* **42**, 5945–5954 (2003).
20. Harmon, F. G. & Kowalczykowski, S. C. RecQ helicase, in concert with RecA and SSB proteins, initiates and disrupts DNA recombination. *Genes Dev.* **12**, 1134–1144 (1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to members of the laboratory for their comments on this work. A.L.F. was funded by an American Cancer Society Postdoctoral Fellowship (PF-08-046-01-GMC) and S.C.K. was supported by the National Institutes of Health (GM-62653 and GM-64745).

Author Contributions A.L.F. and S.C.K. conceived the general ideas, designed the experiments and interpreted the data. A.L.F. performed experiments. A.L.F. and S.C.K. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.C.K. (sckowalczykowski@ucdavis.edu).

METHODS

Microscope. The instrument that was developed was based on an Eclipse TE2000-U inverted microscope with a total internal reflected fluorescence (TIRF) attachment (Nikon) using a CFI Plan Apo TIRF 100 \times , 1.45 numerical aperture oil-immersed objective. Infrared laser trapping, operated in position-clamp mode, was achieved almost exactly as previously described²¹ with the addition of a polarizer (Newport) to split the beam and generate two traps, and a steering mirror (Newport) to control the x - y position of one of the beams. Fluorescence of the sample in TIRF mode was achieved by excitation using a Cyan 488-nm laser (Picoarro) or a 561-nm laser (Cobolt). Epifluorescence illumination was achieved with an X-Cite 120-W mercury vapour lamp (Lumen Dynamics). The fluorescence emission was directed through a polychroic mirror (centre wavelength 515 nm, bandwidth 30 nm; and centre wavelength 600, bandwidth 40 nm; Chroma). Light was guided into a Dual-View apparatus (Optical Insights) where the green and red components were spatially separated (dichroic 565dxc, emission HQ515/30 nm and HQ600/40 nm, Chroma). Movies were captured on a DU-897E iXon CCD camera (Andor, 100-ms exposure) and processed using IQ imaging software (Andor).

Biotinylated λ duplex DNA. Multiple biotin moieties were incorporated into both ends of bacteriophage λ DNA (NEB) by an end-filling reaction. A 30- μ l reaction contained 1 \times NEB buffer number 2, 33 μ M each of dATP, dTTP, dCTP and biotin-11-dGTP (Perkin Elmer), 5 μ g λ DNA and 5 units of Klenow exo⁻ (NEB). The reaction was incubated for 15 min at 25 $^{\circ}$ C then terminated by the addition EDTA to a final concentration of 10 mM and heat inactivation of Klenow at 75 $^{\circ}$ C for 20 min. The reaction was then diluted to a 100- μ l final volume with Nanopure water (Millipore) and passed through an S-400 spin column (GE Healthcare) equilibrated with TE buffer (10 mM Tris-HCl (pH 7.5) and 1 mM EDTA).

Fluorescent ssDNA substrates. DNA primer sequences that were used to amplify defined regions of λ DNA by PCR are the following, for: an 87-bp product for D-loop assay with pUC19 supercoiled DNA: forward primer 5'-biotin-CGACGGCCAGTGAATTCCTCCGA-3', reverse primer 5'-TTACGCCAAGCTTACTCGGGAAACAT-3'; a 162-bp product (identical to λ DNA between base pairs 12,368–12,529): forward primer 5'-biotin-TAACGTCATGTCAGAGCAGAAAAAG-3', reverse primer 5'-GCAATACCATCAAAGGTCTGCGTG-3'; a 430-bp product (identical to λ DNA between base pairs 23,788–24,217): forward primer 5'-biotin-ACTGTTCTTTCGCGTTTGAGG-3', reverse primer 5'-CTATCGGAAGTTCACCGCCAG-3'; and a 1,762-bp product (identical to λ DNA between base pairs 13,767–15,528): forward primer 5'-biotin-GGATGCGGTGAACCTCGTCAAC-3', reverse primer 5'-CCCCTTACTGCTTCTTTACCC-3'.

PCR reactions contained 1 \times ThermoPol buffer (NEB), 0.2 mM dATP, 0.2 mM dCTP, 0.2 mM dGTP, 0.1 mM dTTP, 0.2 mM 5-(3-aminoallyl) dUTP (Fermentas), 0.25 μ g μ l⁻¹ λ DNA (NEB) (pUC19 for a 87-nucleotide substrate), 0.5 μ M each primer and 0.05 U μ l⁻¹ Vent exo⁻ polymerase (NEB). The thermocycler (iCycler, Bio-Rad) program involved initial denaturation at 95 $^{\circ}$ C for 2 min, 30 cycles of a denaturation phase at 95 $^{\circ}$ C for 30 s, an annealing phase at 60.6, 63, 62.2 or 59.4 $^{\circ}$ C for 30 s, for 87-, 162-, 430- or 1762-nucleotide products, respectively, and an extension phase at 72 $^{\circ}$ C for 0.25, 0.25, 1 and 5 min for 87-, 162-, 430- and 1762-nucleotide products, respectively. The final PCR step was extension at 72 $^{\circ}$ C for 5 min. The reactions were then processed with a QIAquick PCR purification kit (Qiagen). Following purification, the DNA was ethanol-precipitated at -20 $^{\circ}$ C. To fluorescently label the PCR products, a 20- μ l reaction containing 10–20 μ g of PCR-generated DNA containing amine-modified nucleotides, 200 mM sodium bicarbonate (pH 9.0) and 5 mM ATTO565 NHS-ester (ATTO-TEC GmbH) was incubated for 1–2 h at 25 $^{\circ}$ C while protected from light. Alexa Fluor 488 succinimidyl ester (Invitrogen) was used to label the 87-nucleotide substrate used in the D-loop assay. Following incubation, 180 μ l Nanopure water was added and a QIAquick PCR purification kit (Qiagen) was used to remove free label. Purified labelled DNA was stored at 4 $^{\circ}$ C until the strand-separation step. Alkali denaturation in combination with the single 5'-biotin incorporated from the forward primer in the PCR reaction was used to produce ssDNA from the fluorescently labelled duplex PCR product as follows: 800 μ l avidin-agarose (400 μ l settled gel; Thermo Scientific) was prepared in a 1.5-ml Eppendorf tube using centrifugation to pellet agarose. All centrifugation steps were performed using a bench-top centrifuge at 4,524g for 1 min. The resin was pelleted and washed three times with 1 ml binding and wash buffer (10 mM Tris-HCl (pH 7.5), 1 mM EDTA and 150 mM NaCl). Fluorescently labelled biotinylated dsDNA (~10–20 μ g, from the PCR reaction above) was diluted to 1 ml with binding and wash buffer. The diluted DNA was added to the prepared avidin-agarose, and mixed end-over-end for 1 h while protected from light. The agarose and bound DNA were pelleted by centrifugation and washed three times with 1 ml binding and wash buffer to remove unbound DNA. The ssDNA was eluted by alkali denaturation of the dsDNA, by addition of 200 μ l of 0.15 M NaOH to the pelleted agarose and mixing

end-over-end for 10 min to release the non-biotinylated strand. The slurry was transferred to an empty micro-spin column (Bio-Rad) and centrifuged at 4,700g to recover the eluted ssDNA. A Microspin S-400 column (GE Healthcare) was used to exchange the ssDNA into the TE buffer. Samples of each fraction were analysed by polyacrylamide or agarose gel electrophoresis. Fractions containing ssDNA were pooled, purified and concentrated with QIAquick PCR purification kit (Qiagen). The DNA concentration was determined using an extinction coefficient of 8,919 M⁻¹ cm⁻¹ at 260 nm, taking into account a correction factor of 0.34 for absorbance at 260 nm by the dye. The dye concentration was determined using an extinction coefficient of 120,000 M⁻¹ cm⁻¹ at 563 nm.

Flowcell fabrication. Channels and holes were etched by CO₂ laser into glass slides (Fisher Scientific 25 \times 75 \times 1 mm) covered with an adhesive abrasive blasting mask (Epilog) using a 30 W Mini-24 Laser Engraver (Epilog Lasers). Following the engraving step, the slides were blasted using 220 grit silicon carbide (Electro Abrasives) to remove residual laser-ablated glass from the channels. A cover glass (Corning No. 1, 24 \times 60 mm) was attached with ultraviolet Optical adhesive number 74 (Norland Products) applied through capillary action. The adhesive was cured by placing the flowcell 30 cm from a 100 W HBO lamp (Zeiss) for 20 min followed by a final heat curing at 70 $^{\circ}$ C for 12 h. PEEK tubing with 0.5 mm inner diameter (Upchurch Scientific) was inserted into each of the etched holes to create inlet and outlet connection ports using 5 min Epoxy (Devcon).

Surface preparation of single-channel flowcell for TIRFM experiments. The surface modification procedure was done at 25 $^{\circ}$ C. The flowcells were cleaned with 1 M NaOH for 30–60 min, and washed twice with 1 ml Nanopure water and then with 1 ml of buffer (25 mM Tris-OAc (pH 7.5), 50 mM NaCl). 1 mg ml⁻¹ biotinylated BSA (Thermo Scientific) in the above buffer was then incubated in the flowcell for 5 min and then washed with 1 ml of buffer. After this, 0.1 mg ml⁻¹ streptavidin (Promega) in buffer was incubated in the flowcell for 5 min then washed with 1 ml of buffer. Finally, the flowcell was blocked with 1.5 mg ml⁻¹ Roche Blocking Reagent (Roche) in buffer for 30–60 min and washed with 1 ml buffer. The prepared flowcell was then mounted on the microscope and attached to the syringe pump (KD Scientific).

D-loop assay. RecA and SSB were purified as previously described^{19,20}. The AlexaFluor 488-labelled 87-nucleotide ssDNA substrate was prepared as described above. A 10- μ l reaction containing 25 mM Tris-HCl (pH 7.5), 10 mM MgCl₂, 1 mM DTT, 2 mM ATP γ S, 100 μ g ml⁻¹ BSA, 4.5 μ M RecA and 105 nM fluorescently labelled 87-nucleotide ssDNA was incubated for 8 min at 37 $^{\circ}$ C. The reaction was started with the addition of 35 nM supercoiled DNA (pUC19) and incubated at 37 $^{\circ}$ C for 20 min. The reaction was stopped by mixing with 5 μ l of stop solution (4.8% SDS, 7 mg ml⁻¹ proteinase K) and incubating for 10 min at 37 $^{\circ}$ C. Products were resolved by electrophoresis in a 1% ultrapure agarose gel (Invitrogen) using TAE (40 mM Tris, 20 mM acetic acid and 1 mM EDTA) at 100 V until the bromophenol blue had migrated 4 cm. The gel was imaged and analysed with a STORM scanner and Image Quant software (Molecular Dynamics). The efficiency of the reaction was calculated as the fraction of ssDNA that formed D-loops multiplied by three to correct for the threefold molar excess of ssDNA relative to supercoiled pUC19 in the reaction.

Single-molecule DNA pairing experiments. Nucleoprotein filaments were formed essentially as described previously⁴ in SM buffer (25 mM Tris-OAc (pH 7.5), 1 mM DTT and 4 mM Mg(OAc)₂); SSB (at a ratio of 1 SSB monomer to 11 nucleotides), 2 nM molecules fluorescent ssDNA and 1 mM ATP γ S were incubated for 10 min at 37 $^{\circ}$ C. RecA was added at a ratio of 1 monomer to 1.7 nucleotides and incubated for 1 h. Nucleoprotein filaments were then diluted tenfold to a final concentration of 0.2 nM in buffer before introduction into the flowcell. In the DNA pairing experiments using TIRFM, biotinylated λ DNA (1 pM, molecules) in SM2 buffer (SM and 50 mM NaCl) was introduced into the flowcell and allowed to bind for several minutes. The flowcell was then washed with 500 μ l SM2 buffer to remove free DNA as well as to extend and attach the second end of the λ DNA molecules. The reaction was started by the addition of 0.2 nM nucleoprotein filaments in SM2 buffer. For ensemble experiments visualized by TIRFM, the nucleoprotein filaments and λ DNA were incubated for 1 h (162-nucleotide substrate) or 30 min (430 nucleotide substrate) at 37 $^{\circ}$ C before visualization in a single-channel flowcell.

Visualization of RecA-mediated pairing with individual DNA dumbbells was performed at 37 $^{\circ}$ C. The flowcell surface was treated for 1 h with BSA (1 mg ml⁻¹) in single-molecule (SM3) buffer (50 mM Tris-OAc (pH 8.2), 50 mM DTT, 1 mM Mg(OAc)₂ and 15% sucrose). Biotinylated λ DNA and buffers were pumped at a linear flow rate of ~100 μ m s⁻¹ into the flowcell. The channels contained SM3 buffer, 18 fM streptavidin-coated polystyrene beads (1 μ m; Bangs Laboratories) and 5 nM YOYO-1 (Invitrogen) (Fig. 2B, a); SM3 buffer, 100 nM YOYO1, and 10 pM (molecules) biotinylated λ DNA (Fig. 2B, b); SM3 buffer (Fig. 2B, c); SM buffer and 15% sucrose (Fig. 2B, d, f). The reaction reservoir contained 0.2 nM nucleoprotein filaments in SM with 15% sucrose and 0.5 mM ATP γ S (Fig. 2B, e).

Data analysis. Data were analysed using GraphPad Prism v5.04. The kinetic data were fit to a single exponential function ($Y = Y_0 + (\text{Plateau} - Y_0)(1 - e^{-kx})$). In Fig. 4b, c, the time courses do not pass through the origin. We are not certain whether this is an intrinsic characteristic of the homology search or a limitation of the experimental procedure: for example, the time for the DNA to relax from flow-induced stretching after movement into the reservoir. We note that the half-time for the relaxation of extended λ DNA is ~ 6 s (ref. 22); during this time the

dsDNA is not in its equilibrium coiled configuration and initial interaction with the RecA nucleoprotein filament would be limited by the DNA polymer dynamics.

21. Bianco, P. R. *et al.* Processive translocation and DNA unwinding by individual RecBCD enzyme molecules. *Nature* **409**, 374–378 (2001).
22. Perkins, T. T., Quake, S. R., Smith, D. E. & Chu, S. Relaxation of a single DNA molecule observed by optical microscopy. *Science* **264**, 822–826 (1994).

CAREERS

UNITED STATES Biomedicine, genomics and energy get charitable donations **p.431**

CHILE Tax credit may boost investment in mining, forestry and agriculture **p.431**

NATUREJOBS For the latest career listings and advice www.naturejobs.com

A SCHEIN/CORBIS



FUNDING

Got to get a grant

A great idea will get applicants only so far. But there are other strategies that can add to the chances of success.

BY KAREN KAPLAN

With his primary grant coming to an end, neuroscientist Thomas Mrsic-Flogel was more than a little stressed. He had launched his lab at University College London (UCL) with a career-development fellowship from the Wellcome Trust in London, but it was set to expire by mid-2011. In 2010, with a worldwide recession in full swing, Mrsic-Flogel knew that he was hardly guaranteed to land a new grant.

He decided to apply for another Wellcome

fellowship, proposing a project on how neuronal networks process visual stimuli. Applications had a discouraging success rate of about 20%, but the grant could be renewed every five years, which Mrsic-Flogel found attractive. He won the award — a £1.7-million (US\$2.7-million) senior research fellowship, which pays his salary and lets him purchase lab equipment and support a couple of graduate research associates.

Mrsic-Flogel attributes his success to more than luck. He followed the application guidelines to the letter, making sure that his proposal was both high-impact and innovative. He spent

a year preparing it, including developing his idea and gathering preliminary data. And he sought input from dozens of people, from UCL grant advisers to colleagues in neuroscience and other fields, in effect creating an informal peer-review panel. He revised the document several times, once deleting an entire section, and when something stumped him, Mrsic-Flogel called grant recipients he knew to find out how they had dealt with similar problems.

In the current funding environment, the odds of winning a grant or fellowship are very slim. But Mrsic-Flogel's success demonstrates some helpful strategies and guidelines — articulating an original idea, seeking feedback from multiple sources and writing concisely — for putting together a winning proposal.

EXCELLENT SCIENCE

Before all else, applicants must make sure that they are presenting excellent and original science, say grant programme officers and successful applicants. “You should be proposing a novel kind of research — not just ►

► continuing some standard research you're already doing," says Jochen Wosnitzer, director of the Dresden High Magnetic Field Laboratory in Germany and chairman of the review board for the German Research Foundation (DFG) in Bonn, the country's main grant-giving agency. To make sure that their projects are innovative, applicants should bounce ideas off colleagues and painstakingly comb through the literature.

Grant officers are generally looking for work that could have an enduring influence. It should inherently lead to further study, although not necessarily to immediate applications. "Have you thought about what happens next?" asks David Crosby, programme manager for the UK Medical Research Council (MRC) in London and Swindon. Research funded by the MRC doesn't have to lead to a disease cure in three years, "but you do need to think about the implications of your work," he says. "If you're generating a fundamental insight, what is the consequence of that? How does that help the whole field? How might it go on to be utilized? How will it impact the science community and the public at large?" Researchers applying for a grant from a multinational organization, such as the European Commission's Marie Curie Actions, or funding from the European Molecular Biology Organization (EMBO) in Heidelberg, Germany, will also need to explain how their proposal would have benefits beyond their own country.

Early-career researchers should keep in mind that many granting agencies frown on proposals linked to or associated with work done by the applicant's mentor. "You have to show that you're an independent-thinking scientist taking a different track from your former supervisor," says Gerlind Wallon, deputy director of EMBO and manager of the organization's Young Investigators programme.

There are no hard and fast rules on which funder to approach, say granting and funding agencies. Colleagues with their own grants can offer advice; early-career scientists applying to the US National Institutes of Health (NIH), for example, can get the names of successful grantees from NIH RePORTER (go.nature.com/32v6n5). It can also be extremely helpful to speak directly to the funder; however, programme managers recommend that applicants first learn the agency's remit by closely reading its website and grant materials. "Absolutely come to us," says Crosby. "Phone up the funder

and say, 'I've got this idea that I think pertains to your strategic interest. You've got a highlight notice on your website that says 'systems biology' — what do you mean by that? Does my idea fit into that bracket?'"

Crosby points out that a researcher's institution may also have a preference; for example, the MRC and Wellcome Trust both fund biomedical proposals, but the MRC pays some indirect costs and overheads to the institution that other funders don't, and so might be more attractive.

NUTS AND BOLTS

Applicants must effectively outline their ideas in the application, including a clear and direct hypothesis along with the expected results. Programme managers say that an application for funding to 'explore a cell receptor's signalling mechanisms', for example, is unlikely to be successful because it sounds vague and doesn't seek to prove anything. But a proposal to confirm that a particular protein is involved in a cellular reaction, for example — one that includes preliminary results and explains the potential impact of the discovery — would have a far better chance.

Some applications call for both a summary, aimed at reviewers who are not in the relevant field, and an abstract, for those who are. Most also have a section for a research plan, in which applicants can explain technical details. However, reviewers who see an application for the first (and perhaps only) time in a review-panel meeting usually turn immediately to the summary, say grant officers. That is where applicants should persuasively and succinctly explain exactly why their proposal deserves funding. "It's important to be able to clearly articulate your ideas," says Crosby. "If you can't do that, you're not going to be able to inspire enthusiasm." Some funders also call for a project description or narrative, but veteran grant-writers say that if there is a choice, it is best to make the strongest case in the summary.

Focus is key. If the summary is too technical or rambling, the application's score will suffer, even if the idea itself is brilliant. "A bad summary is really disastrous," says Andrea Hutterer, programme manager for EMBO fellowships. "It sets the tone for how I read the rest of the application."

Applicants must state their research objective clearly and straight away. "The first sentence should begin, 'The research objective of this proposal is ...,'" says George Hazelrigg, a programme officer for design and integration engineering at the US National Science Foundation. "Every inch from the top that I have to go down in the proposal to find this sentence lowers the rating by about one percentage point."

It is wise to get editing and streamlining recommendations from as many senior colleagues as possible, both in and outside the research field, and to check the funder's website for

advice. In a mock application on the NIH website, the 'before' summary, meant to demonstrate pitfalls, is long, rambling and technical ("G-protein over-activation triggers a biochemical signaling cascade that leads to b-AR desensitization and down-regulation ..."), and contains several acronyms. The corrected 'after' summary is clear and direct: "Conges-

tive heart failure is a common and lethal disease in the United States. Current medications ... improve survival in some, but not all, patients. ... This research will enhance our understanding of the cellular and molecular mechanisms underlying sympathetic neuron dysfunction that may progress to heart disease, and may identify a possible novel pharmaceutical target."

Applicants should make sure to request an appropriate amount of funding. Too little and there won't be enough

money to finish the project — and it is next to impossible, say grant officers, to get supplementary funding. Too much and reviewers are likely to question the applicant's competence. "It implies that you don't know what you're doing and don't have a realistic grasp of the project," says Crosby. Applicants can get help with calculations from their department heads, senior supervisors and mentors. For the costs of supplies, such as lab mice, they can talk to the institutional research office.

SWEAT THE SMALL STUFF

Other fundamental requirements may sound mundane or even silly — but failing to adhere to them can derail an application (see 'Grant-writing blunders'). Investigators should read and follow all application instructions carefully: most stipulate length and format, including particular typefaces, fonts, font sizes and margins. It does not pay to deviate from these in the hope of cramming in more text or figures, warn programme managers.

"Bend over backwards to give us what we want," advises Maryrose Franko, senior programme officer for graduate science education at the Howard Hughes Medical Institute in Chevy Chase, Maryland. Reviewers don't want to sift through an application to find an investigator's most significant published work or squint to read the text, she says. "If we say 12-point font and you give us 10, the reviewers don't even want to look at it."



"It's important to be able to clearly articulate your ideas. If you can't do that, you're not going to be able to inspire enthusiasm."

David Crosby



"Bend over backwards to give us what we want."

Maryrose Franko

MRC

C. VARGAS/HIMI

DOS AND DON'TS

Grant-writing blunders

- Avoid being too ambitious — don't propose a study that would take decades. Grant officers can tell when an applicant is overextending.
- Don't use abbreviations, acronyms, jargon or highly technical language. Reviewers who aren't familiar with your field will get annoyed and may think that you are trying to cover up for a lack of knowledge or ability to carry out the experiment.
- Don't give short shrift to explaining why your proposal is important. Reviewers

don't already know. Explain the study's impact, advances and potential.

- Make the application easy to read — don't cram it with text, use too-small fonts or miniaturize any figures.
- Get lots of colleagues from within and outside your field to review your application closely and provide written responses.
- Make sure that you're asking for an appropriate sum. If you request too much or too little, reviewers will conclude that you don't know what you're doing. **K.K.**

Proposals must be easy to read, agree Stephen Russell and David Morrison, co-founders of Grant Writers' Seminars and Workshops, a consulting business in Los Olivos, California, that helps clients with applications. "Reviewers read grant applications for only one reason — because they have to," says Russell. To help them, he and Morrison recommend making margins wider than the minimum, using an easy-to-read typeface and font size such as 12-point Arial — or whatever is specified in the instructions — and adding spaces between paragraphs and sections.

Spelling errors and poor grammar may not immediately disqualify an application, but they could lower the score, or at the very least give a bad impression. "Bad English and typos are an annoyance factor that reviewers have to overcome," says Wallon. "If it's done sloppily, I wouldn't recommend it."

But scientists don't necessarily need to hire a consultant to make sure that their application is letter-perfect, say programme managers. "Using a commercial consultant gives your application a tone that panel members will detect. We're looking for a contribution from the individual," says Alex Martin Hobdey, head of the unit for starting grants at the European Research Council in Brussels. Consultant-assisted applications tend to sound too slick or smooth — it is more effective to get editing recommendations from colleagues.

Submissions that are incomplete or past deadline are certain to be disqualified. Hutterer says that out of the 850 applications to EMBO's fellowship programme each year, some 150 are unfinished and thus immediately ineligible. And Dennis Abbott, a spokesman for the Marie Curie Actions programme, decries late submissions. "No matter how good your application is, it's too late," he says. "Deadlines are set for a reason."

SHADES OF EXCITEMENT

Applicants need to communicate the pay-offs of the research straight away. Russell says that a common mistake is to write a title that could be reused for future renewal applications. For example, he says, 'Studies of renal disease' is accurate but generic. He suggests evoking a salient image or concept — something more like 'Contribution of anti-idiotypic antibodies to pathogenesis of acute glomerulonephritis'. He warns applicants not to let snappiness obscure the content of the proposal — something like 'Breakthrough treatment strategies to cure acute glomerulonephritis' draws attention but is sensationalistic and vague.

It helps to be positive and enthusiastic in project summaries, abstracts and research questions — but to include a back-up plan. "You need to say that you expect that this approach will work; however, if it doesn't, you will be prepared to do this and this," says Morrison. "It's all about asserting confidence in your ability to do this research, backed up by your fallback of alternative strategies."

Ultimately, once the mechanics are right, it boils down to convincing reviewers that the application deserves funding. "If you can't convey your excitement and the importance of your proposal and what you think your results will be," says Franko, "then you're not going to get good scores." ■

Karen Kaplan is Nature's assistant Careers editor.



"You have to show that you're an independent-thinking scientist taking a different track."

Gerlind Wallon

UNITED STATES

Charity supports science

At least 10 of the top 50 US charitable donors of 2011 gave funds to support scientific research, according to the *Philanthropy 50* report released on 6 February by *The Chronicle of Philanthropy* in Washington DC. The top 50 donors gave a total of US\$10.4 billion, up from \$3.3 billion in 2010. The *Chronicle* speculates that the increase is due to some economic recovery and a perceived need for funds at universities. Donations included \$70 million to the Allen Institute for Brain Science in Seattle, Washington, for neuroscience and genomics research; \$59.2 million to the Ellison Medical Foundation in Bethesda, Maryland, for biomedical research; and \$25 million to Yale University in New Haven, Connecticut, to launch an energy-research institute.

CHILE

Tax credit for research

The Chilean government hopes that a tax incentive will boost investment in research and development (R&D), and create jobs. The scheme triples the maximum tax credit for research-investment costs; eliminates a 15% tax on gross sales, easing the financial burden for entrepreneurs and start-ups; and can offset costs related to securing intellectual-property rights. The law will come into effect this year. Pablo Longueira, Chile's economics minister, expects companies in mining, forestry, energy, agriculture and aquaculture to expand their research. "We believe that many of the new PhDs that are currently being trained outside of the country will return to work for R&D projects under this new law," he says.

ANIMAL HEALTH

Allen school expanding

Recruitment has begun at Washington State University's Paul G. Allen School for Global Animal Health in Pullman, where a new research facility will open in May. By 2015, administrators hope to hire 13 researchers to detect emerging cross-species diseases, develop vaccines and work on transmission control, says director Guy Palmer. Hiring is supported by US\$51 million in donations from Microsoft co-founder Paul Allen and the Bill & Melinda Gates Foundation in Seattle, Washington; another \$14 million is earmarked for programmes including training students in East Africa.

DOS AND DON'TS

Grant-writing blunders

- Avoid being too ambitious — don't propose a study that would take decades. Grant officers can tell when an applicant is overextending.
- Don't use abbreviations, acronyms, jargon or highly technical language. Reviewers who aren't familiar with your field will get annoyed and may think that you are trying to cover up for a lack of knowledge or ability to carry out the experiment.
- Don't give short shrift to explaining why your proposal is important. Reviewers

don't already know. Explain the study's impact, advances and potential.

- Make the application easy to read — don't cram it with text, use too-small fonts or miniaturize any figures.
- Get lots of colleagues from within and outside your field to review your application closely and provide written responses.
- Make sure that you're asking for an appropriate sum. If you request too much or too little, reviewers will conclude that you don't know what you're doing. **K.K.**

Proposals must be easy to read, agree Stephen Russell and David Morrison, co-founders of Grant Writers' Seminars and Workshops, a consulting business in Los Olivos, California, that helps clients with applications. "Reviewers read grant applications for only one reason — because they have to," says Russell. To help them, he and Morrison recommend making margins wider than the minimum, using an easy-to-read typeface and font size such as 12-point Arial — or whatever is specified in the instructions — and adding spaces between paragraphs and sections.

Spelling errors and poor grammar may not immediately disqualify an application, but they could lower the score, or at the very least give a bad impression. "Bad English and typos are an annoyance factor that reviewers have to overcome," says Wallon. "If it's done sloppily, I wouldn't recommend it."

But scientists don't necessarily need to hire a consultant to make sure that their application is letter-perfect, say programme managers. "Using a commercial consultant gives your application a tone that panel members will detect. We're looking for a contribution from the individual," says Alex Martin Hobdley, head of the unit for starting grants at the European Research Council in Brussels. Consultant-assisted applications tend to sound too slick or smooth — it is more effective to get editing recommendations from colleagues.

Submissions that are incomplete or past deadline are certain to be disqualified. Hutterer says that out of the 850 applications to EMBO's fellowship programme each year, some 150 are unfinished and thus immediately ineligible. And Dennis Abbott, a spokesman for the Marie Curie Actions programme, decries late submissions. "No matter how good your application is, it's too late," he says. "Deadlines are set for a reason."

SHADES OF EXCITEMENT

Applicants need to communicate the pay-offs of the research straight away. Russell says that a common mistake is to write a title that could be reused for future renewal applications. For example, he says, 'Studies of renal disease' is accurate but generic. He suggests evoking a salient image or concept — something more like 'Contribution of anti-idiotypic antibodies to pathogenesis of acute glomerulonephritis'. He warns applicants not to let snappiness obscure the content of the proposal — something like 'Breakthrough treatment strategies to cure acute glomerulonephritis' draws attention but is sensationalistic and vague.

It helps to be positive and enthusiastic in project summaries, abstracts and research questions — but to include a back-up plan. "You need to say that you expect that this approach will work; however, if it doesn't, you will be prepared to do this and this," says Morrison. "It's all about asserting confidence in your ability to do this research, backed up by your fallback of alternative strategies."

Ultimately, once the mechanics are right, it boils down to convincing reviewers that the application deserves funding. "If you can't convey your excitement and the importance of your proposal and what you think your results will be," says Franko, "then you're not going to get good scores." ■

Karen Kaplan is Nature's assistant Careers editor.



"You have to show that you're an independent-thinking scientist taking a different track."

Gerlind Wallon

UNITED STATES

Charity supports science

At least 10 of the top 50 US charitable donors of 2011 gave funds to support scientific research, according to the *Philanthropy 50* report released on 6 February by *The Chronicle of Philanthropy* in Washington DC. The top 50 donors gave a total of US\$10.4 billion, up from \$3.3 billion in 2010. The *Chronicle* speculates that the increase is due to some economic recovery and a perceived need for funds at universities. Donations included \$70 million to the Allen Institute for Brain Science in Seattle, Washington, for neuroscience and genomics research; \$59.2 million to the Ellison Medical Foundation in Bethesda, Maryland, for biomedical research; and \$25 million to Yale University in New Haven, Connecticut, to launch an energy-research institute.

CHILE

Tax credit for research

The Chilean government hopes that a tax incentive will boost investment in research and development (R&D), and create jobs. The scheme triples the maximum tax credit for research-investment costs; eliminates a 15% tax on gross sales, easing the financial burden for entrepreneurs and start-ups; and can offset costs related to securing intellectual-property rights. The law will come into effect this year. Pablo Longueira, Chile's economics minister, expects companies in mining, forestry, energy, agriculture and aquaculture to expand their research. "We believe that many of the new PhDs that are currently being trained outside of the country will return to work for R&D projects under this new law," he says.

ANIMAL HEALTH

Allen school expanding

Recruitment has begun at Washington State University's Paul G. Allen School for Global Animal Health in Pullman, where a new research facility will open in May. By 2015, administrators hope to hire 13 researchers to detect emerging cross-species diseases, develop vaccines and work on transmission control, says director Guy Palmer. Hiring is supported by US\$51 million in donations from Microsoft co-founder Paul Allen and the Bill & Melinda Gates Foundation in Seattle, Washington; another \$14 million is earmarked for programmes including training students in East Africa.

PICNIC WITH ANTS

Come fly with me.

BY MARK W. MOFFETT

I'm Gerry Blandsides and this is recording tape 18 for my postdoctoral grant 978-2023, *The Geopolitical Significance of the Ants in Namibian Mud Wallows: An Über-synthesis*. It's day nine of my observations on the *Pheidole* 'big-headed' ant colony nesting east of Otjiwarongo in warthog splash puddle and latrine 62Ω.

07:00 71 ants passed by along their foraging trail in the last minute. Their diet remains unclear, though as before they transport bits of blackened material.

07:02 No change.

7:04 Nothin'.

7:06 Nada.

7:08 Yawn.

7:10 Zzz.

7:12 Kill me.

07:14 I count 69 ants this last minute. An occasional ant hauls a pebble to one of the rings of sand they've let accumulate near the trail border. (Note to self: this pointless activity makes me imagine that the ants are as bored as I. *C'est la vie!* For ultimately, I shall triumph with my unifying concept of post-Jurassic hypertrophy of ant neosocial meta-structure under conditions of intense swine excreta interactivity and its effects on antennal waving dynamics.)

07:16 73 ants per minute. One ant stops, grooms itself. Fly lands on my nose.

07:18 67 ants. (Note: trail usage remains stable, contrary to the conjectures of those bohemian Yale intellectuals with their fancy graph paper. At least this prediction from Appendix 142 of my grant will be borne out? Finishing my count, I give the grooming ant a cheery salute as it turns in my direction.)

07:20 I slap the fly. The ant stops grooming, runs to the moribund bug. It pauses to look at me, then looks at the fly, then at me again, before scurrying off.

07:22 Six ants gather around the fly, start a small fire under it within one of those rings of sand they had deposited earlier. (Note: fly wings crackle as they burn.) The team rotates the fly so that it roasts evenly, turning from a golden dipterous brown

to a deep smoky grey. (Another note: isn't cooking unique to humans? Follow-up grant assured if this finding is reproducible.)

07:24 The charbroiled scent is driving me crazy. The ants have taken herbs from nearby shrubs, adding the redolence of oregano, but with more vibrant undertones! I pull a sandwich from my field vest, but the baloney disappoints.



07:26 Ants remove fly from fire, carve the smoked meat. (Note: salivating! And if I recall correctly, insects have less fat and more protein than steak.)

07:28 The ant that had been grooming itself — and I'm *certain* it is the same individual — has returned with six others. (Note to self: as with other *Pheidole* ant species, this recruitment of assistance was doubtless accomplished by use of chemical scent that the leader ant releases into the air. The authority on such pheromones is Professor E. O. Wilson — might he support a field-work project on this?) The newly arrived ant workers lug half the butchered fly in my direction, then they gaze up at me and back away slowly.

Tastes delicious, I knew it would.

[Long interval of static on recording tape.] ... need a seductive marketing name, like the ones restaurants give ugly game fish.

Arthrofowl? Miniqual? Souperfly? Kosher-bug? Chicken Little? MicroMcNuggats? Flying lobster?

17:56 313 ants/minute. 87 little bonfires flicker at my feet in the last glow of sunset. A steady supply of swatted flies — now including crepuscular mosquitoes of the genus *Anopheles* — keeps my ant colony busy. Their recipe improves each time I squash an ant who overcooks. (Note: natural selection in action? I should write a grant on this topic also. Then again, forget the science — boring, boring! Ants that cook, who cares? Future is assured if I replicate this — succulent? aphrodisiant? antbrosia? — recipe without them.)

17:58 The 50/50 split that the ants are giving me seems fair. But still, I calculate I will require thousands of flies to maintain the diet. What must it be like for an elephant to depend on peanuts, handed out one at a time by children? So no more bug repellent for me! Let Bugs Come Hither. Anyway, DEET gives each morsel a decided bitterness — makes the ants queasy as well.

Hold on. More ants are watching me. It seems a fly has landed on my ear, and now there's a mosquito on my forehead. Back in two minutes.

[Another interval of indecipherable static.]

19:22 Too many ants are arriving to count — 4,000 a minute? No matter how many flies I swat, I can no longer keep up with their needs. Worse, the flames below me have merged into a single conflagration that is singing my hair. My face is blistering; eyes water from the heat. (Note: Stupid ants, how can they cook anything now? Setbacks like this could cause delays if I decide to approach the Food Network. Wait, look at *that!* The ants swarm my legs, some of them carrying herbs. It's a recruitment response a thousand times more intense than they show to a fly! Is a different pheromone involved? I will take copious notes, but first I must figure out why I can't move.)

[Recording ends.] ■

Mark W. Moffett is a Smithsonian entomologist who has won the Lowell Thomas Medal from the Explorers Club and the Bowdoin Medal for writing from Harvard. After completing a PhD under E. O. Wilson, he spent years watching ants for his recent book, *Adventures Among Ants*.

➤ NATURE.COM
Follow Futures on
Facebook at:
go.nature.com/mtoodm

Isotope fractionation in silicate melts

ARISING FROM G. Dominguez, G. Wilkins & M. H. Thieme *Nature* **473**, 70–73 (2011)

Experiments show that temperature gradients in silicate melts lead to isotope fractionation, where the heavier isotopes concentrate in cold regions and light isotopes concentrate in hot regions^{1–4}. Dominguez *et al.*⁵ present a phenomenological model based on quantum effects that provides a good fit to these experimental results, and argue that “consideration of the quantum mechanical zero-point energy of diffusing species is essential for understanding diffusion at the isotopic level”. However, we point out that the zero-point energy required to fit their model to experimental results is unphysically large, and that isotopic fractionation similar to that observed in silicate melts is found in systems where quantum effects are absent. Therefore, the conclusion that quantum effects underlie isotope fractionation in silicate melts with temperature gradients is not justified.

To fit experimental data, the Dominguez *et al.*⁵ model requires a zero-point energy (ZPE) for ²⁶Mg of ~0.4 eV. The atomic motion giving rise to the ZPE is vibrational, and can be modelled by a harmonic oscillator for which $ZPE = (1/2)h\nu$, where h is Planck’s constant and ν is the vibrational frequency. (Here for convenience we consider $\tilde{\nu} = \nu/c$, where c is the velocity of light.) The value $ZPE \approx 0.4$ eV corresponds to $\tilde{\nu} \approx 6,500$ cm^{−1}, which is much larger than the highest vibrational frequencies (~1,300 cm^{−1}) observed in anhydrous silicate melts⁶. In fact, $\tilde{\nu} \approx 6,500$ cm^{−1} is larger than the vibrational frequency in any material whatsoever (the highest vibrational frequency we are aware of is that for H₂, where $\tilde{\nu} \approx 4,395$ cm^{−1})⁷. Thus a ZPE of ~0.4 eV is not physically relevant.

The unphysically large ZPE in the model of Dominguez *et al.*⁵ leads to predictions of relative diffusivities of isotopes that are in poor agreement with experiments. For example, their model (equations (11) and (12), and $ZPE(^{26}\text{Mg}) = 0.4$ eV) predicts $D(^{24}\text{Mg})/D(^{26}\text{Mg}) = 1.13$ at 1,500 K. In contrast, experiments on silicate melts find $D(^{24}\text{Mg})/D(^{26}\text{Mg}) = 1.004$ (ref. 2). Thus, the Dominguez *et al.* model predicts an isotope effect for relative diffusivities that is more than 30 times larger than found experimentally (13% versus 0.4%).

Finally, we note that isotope fractionation in temperature gradients occurs in systems where quantum effects are not relevant; this implies that quantum effects are not a necessary condition for isotope fractionation to occur (whereas they are a necessary condition in the Dominguez *et al.*⁵ model). For example, significant fractionation of isotopes is seen in gases held in a temperature gradient^{8–10}. In gases, quantum ZPE (arising from confinement) plays no role because molecules typically are far apart. Thermal fractionation of isotopes is also observed in molecular dynamics simulations of condensed phase systems¹¹ based on classical mechanics—these simulations ignore quantum effects, and in contrast to the model of Dominguez *et al.*⁵ include no phenomenological considerations. In both of these cases, heavier isotopes concentrate in cold

regions and light isotopes concentrate in hot regions, consistent with experimental observations on silicate melts and all other condensed phase systems that have been studied. This effect is understood theoretically in terms of classical mechanics¹², and quantitative agreement is obtained between this theory and experiment¹³.

Daniel J. Lacks¹, James A. Van Orman² & Charles E. Lesher³¹Department of Chemical Engineering, Case Western Reserve University, Cleveland, Ohio 44106, USA.

email: daniel.lacks@case.edu

²Department of Geological Sciences, Case Western Reserve University, Cleveland, Ohio 44106, USA.³Department of Geology, University of California, Davis, California 95616, USA.

Received 13 May; accepted 11 November 2011.

1. Kyser, T. K., Lesher, C. E. & Walker, D. The effects of liquid immiscibility and thermal diffusion on oxygen isotopes in silicate liquids. *Contrib. Mineral. Petrol.* **133**, 373–381 (1998).
2. Richter, F. M., Watson, E. B., Mendybaev, R. A., Teng, F.-Z. & Janney, P. E. Magnesium isotope fractionation in silicate melts by chemical and thermal diffusion. *Geochim. Cosmochim. Acta* **72**, 206–220 (2008).
3. Richter, F. M. *et al.* Isotopic fractionation of the major elements of molten basalt by chemical and thermal diffusion. *Geochim. Cosmochim. Acta* **73**, 4250–4263 (2009).
4. Huang, F. *et al.* Isotope fractionation in silicate melts by thermal diffusion. *Nature* **464**, 396–400 (2010).
5. Dominguez, G., Wilkins, G. & Thieme, M. H. The Soret effect and isotopic fractionation in high-temperature silicate melts. *Nature* **473**, 70–73 (2011).
6. McMillan, P. F., Wolf, G. H. & Poe, B. T. Vibrational spectroscopy of silicate liquids and glasses. *Chem. Geol.* **96**, 351–366 (1992).
7. Karplus, M. & Porter, R. N. *Atoms and Molecules: An Introduction for Students of Physical Chemistry* 369 (Benjamin-Cummings, Menlo Park, California, 1970).
8. Nier, A. O. The concentration of carbon 13 by thermal diffusion. *Phys. Rev.* **57**, 30–34 (1940).
9. Watson, W. W. Thermal separation of isotopes. *Phys. Rev.* **57**, 899–902 (1940).
10. Grachev, A. M. & Severinghaus, J. P. Laboratory determination of thermal diffusion constants for ²⁹N₂/²⁸N₂ in air at temperatures from −60 to 0 °C for reconstruction of magnitudes of abrupt climate changes using the ice core fossil-air paleothermometer. *Geochim. Cosmochim. Acta* **67**, 345–360 (2003).
11. Reith, D. & Müller-Plathe, F. On the nature of thermal diffusion in binary Lennard-Jones liquids. *J. Chem. Phys.* **112**, 2436–2443 (2000).
12. Furry, W. H., Jones, R. C. & Onsager, L. On the theory of isotope separation by thermal diffusion. *Phys. Rev.* **55**, 1083–1095 (1939).
13. Stier, L. G. The coefficients of thermal diffusion of neon and argon and their variation with temperature. *Phys. Rev.* **62**, 548–551 (1942).

Author Contributions All authors contributed to the analysis and the writing of the paper.

Competing Financial Interests Declared none.

doi:10.1038/nature10764

Dominguez *et al.* replyREPLYING TO D. J. Lacks, J. A. Van Orman & C. E. Lesher *Nature* **482**, <http://dx.doi.org/10.1038/nature10764> (2012)

Lacks *et al.*¹ argue that our model of isotopic fractionation in thermal gradients in silicate melts² does not agree with measurements of the ratio of diffusivities seen in silicate melts. This statement is based on an over-interpretation of our model into non-steady-state applications, such as chemical fractionation, because the model we presented

treats the quantized energy levels of the transition state as being equal to each other (the partition function $Z(\text{TS}) = 1$). This was warranted, as our main interest was in finding the steady-state solution to isotopic fractionation in a closed system (which is insensitive to the transition state). The potential importance of the transition state in determining

BRIEF COMMUNICATIONS ARISING

the ratios of diffusivities of He isotopes in a geologic system has previously been noted³, and future work will need to clarify the importance of the transition state for kinetic isotopic fractionation in silicate systems, particularly the ratio of diffusivities.

Lacks *et al.*¹ point out that isotopic fractionation due to temperature gradients in the gas phase have been observed (for example, ref. 4), but they do not provide proof that classical mechanics quantitatively explains these observations. Furthermore, Lacks *et al.*¹ (and references therein) provide no evidence that molecular dynamics simulations reproduce the isotopic fractionations observed in real gas systems, much less high-temperature condensed phases. A full understanding of isotopic fractionation for complex (diatomic, polyatomic) species, even in the gas phase, is likely to require quantum mechanics because of the involvement of quantized vibrations⁵.

No evidence is presented by Lacks *et al.*¹ that molecular dynamics simulations, based on Lennard-Jones interactions, are capable of reproducing the isotopic fractionations of elements and of capturing the strong potential energy interactions that characterize silicate melts and other systems where diffusion is a strong function of temperature. Because of the magnitude of the activation energies involved, only a very small fraction (f) of all particles acquire enough energy to overcome activation energy barriers: $f \approx \exp(-E_a/k_B T)$. Here E_a is activation energy, k_B Boltzmann's constant, and T temperature. For typical activation energies of 2–3 eV, this fraction is rather small ($\sim 10^{-8}$), thus making it difficult for molecular dynamics simulations, which typically employ $\sim 10^3$ particles, to realistically capture the thermodynamics associated with isotopic fractionation in a highly interacting condensed matter system. We suggest caution in interpreting simple binary mixture molecular dynamics simulations that rely on 'unphysical' means of implementing heat transport⁶.

Last, Lacks *et al.*¹ point out that the vibrational frequencies needed to explain the steady-state fractionations are physically unrealistic. We note that this frequency may or may not correspond to a physical frequency, as the initial reactant state consists of three independent vibrational components that we, for the sake of clarity and simplicity, incorporated into one effective frequency (thus the effective frequency would be higher than each of the independent components)⁷. The

infrared vibrational features that are observed in natural systems are sensitive only to differences in energy levels, not the zero-point energy. We also note that, as infrared vibrational features in a silicate melt are overwhelmingly dominated by modes associated with Si–O stretching and bending, it is unclear whether Raman or vibrational infrared spectra include features associated with diffusing interstitial species such as Mg, Fe and Ca.

In short, Lacks *et al.*¹ have provided no new data or physical model that quantitatively explains the empirical observations of steady-state isotopic fractionation in silicate melts. Furthermore, over 50 years of work in isotope effects in physical chemistry and isotope geochemistry support the role of quantum mechanics in isotopic fractionation processes.

G. Dominguez^{1,2}, G. Wilkins³ & M. H. Thiemens²

¹California State University, San Marcos, Department of Physics, California 92096, USA.

email: gdominguez@csusm.edu

²University of California, San Diego, Department of Chemistry and Biochemistry, California 92093, USA.

³University of California, San Diego, Department of Mathematics, California 92093, USA.

1. Lacks, D. J., Van Orman, J. A. & Leshner, C. E. Isotope fractionation in silicate melts. *Nature* **482**, <http://dx.doi.org/10.1038/nature10764> (this issue).
2. Dominguez, G., Wilkins, G. & Thiemens, M. On the Soret effect and isotopic fractionation in high-temperature silicate melts. *Nature* **473**, 70–73 (2011).
3. Trull, T. W. & Kurz, M. D. Isotopic fractionation accompanying helium diffusion in basaltic glass. *J. Mol. Struct.* **485–486**, 555–567 (1999).
4. Grachev, A. M. & Severinghaus, J. P. Laboratory determination of thermal diffusion constants for $^{29}\text{N}_2/^{28}\text{N}_2$ in air at temperatures from -60 to 0°C for reconstruction of magnitudes of abrupt climate changes using the ice core fossil-air paleothermometer. *Geochim. Cosmochim. Acta* **67**, 345–360 (2003).
5. Saxena, S. C. & Mathur, B. P. Thermal diffusion in isotopic gas mixtures and intermolecular forces. *Rev. Mod. Phys.* **38**, 380–390 (1966).
6. Reith, D. & Florian, M.-P. On the nature of thermal diffusion in binary Lennard-Jones liquids. *J. Chem. Phys.* **112**, 2436–2453 (2000).
7. Vineyard, G. H. Frequency factors and isotope effects in solid state rate processes. *J. Phys. Chem. Solids* **3**, 121–127 (1957).

doi:10.1038/nature10765

Abrupt acceleration of a ‘cold’ ultrarelativistic wind from the Crab pulsar

F. A. Aharonian^{1,2}, S. V. Bogovalov³ & D. Khangulyan⁴

Pulsars are thought to eject electron–positron winds that energize the surrounding environment, with the formation of a pulsar wind nebula¹. The pulsar wind originates close to the light cylinder, the surface at which the pulsar co-rotation velocity equals the speed of light, and carries away much of the rotational energy lost by the pulsar. Initially the wind is dominated by electromagnetic energy (Poynting flux) but later this is converted to the kinetic energy of bulk motion². It is unclear exactly where this takes place and to what speed the wind is accelerated. Although some preferred models imply a gradual acceleration over the entire distance from the magnetosphere to the point at which the wind terminates^{3,4}, a rapid acceleration close to the light cylinder cannot be excluded^{5,6}. Here we report that the recent observations of pulsed, very high-energy γ -ray emission from the Crab pulsar^{7–9} are explained by the presence of a cold (in the sense of the low energy of the electrons in the frame of the moving plasma) ultrarelativistic wind dominated by kinetic energy. The conversion of the Poynting flux to kinetic energy should take place abruptly in the narrow cylindrical zone of radius between 20 and 50 light-cylinder radii centred on the axis of rotation of the pulsar, and should accelerate the wind to a Lorentz factor of $(0.5\text{--}1.0) \times 10^6$. Although the ultrarelativistic nature of the wind does support the general model of pulsars, the requirement of the very high acceleration of the wind in a narrow zone not far from the light cylinder challenges current models.

The Crab pulsar is one of the brightest γ -ray sources in the sky. Both the light curve and the energy spectrum have been studied¹⁰ in great detail by the Large Area Telescope on board NASA’s Fermi Gamma-ray Space Telescope (Fermi). The phase-averaged spectrum is best fitted by a power law with a photon index of $\alpha = 1.97$ and an exponential cut-off at $E_c = 5.8$ GeV (Fig. 1). Although modified ‘outer gap’ models¹¹ do allow an extension of the spectrum up to 10 GeV, the detection of pulsed, very high-energy (VHE) γ -ray emission demands a different radiation component. The extrapolation of the fluxes reported by Fermi to the VHE domain as a power law with photon index $\alpha \approx 3.8$, and the claim that such a formal fit is evidence that γ -rays of gigaelectronvolt (GeV) energies have the same magnetospheric origin as those of teraelectronvolt (TeV) energies^{8,9,12}, in fact requires a drastic revision of basic concepts used at present in magnetospheric models. Moreover, the assumption of a magnetospheric origin for radiation over the entire γ -ray domain contradicts the essentially different light curves reported at GeV (ref. 10) and TeV (refs 7, 9) energies (unless the production sites of these two components are well separated), as well as the apparent tendency of spectral flattening above 100 GeV (Fig. 1).

A natural and more plausible site of production of pulsed VHE γ -rays is the ultrarelativistic wind illuminated by photons originating in the pulsar’s magnetosphere and/or the surface of the neutron star¹³. In the case of the Crab pulsar, the phase-averaged flux of the pulsed (magnetospheric) component exceeds the flux of the thermal emission of the neutron star by two orders of magnitude. The combination of the hard spectral energy distribution of the pulsed emission and the

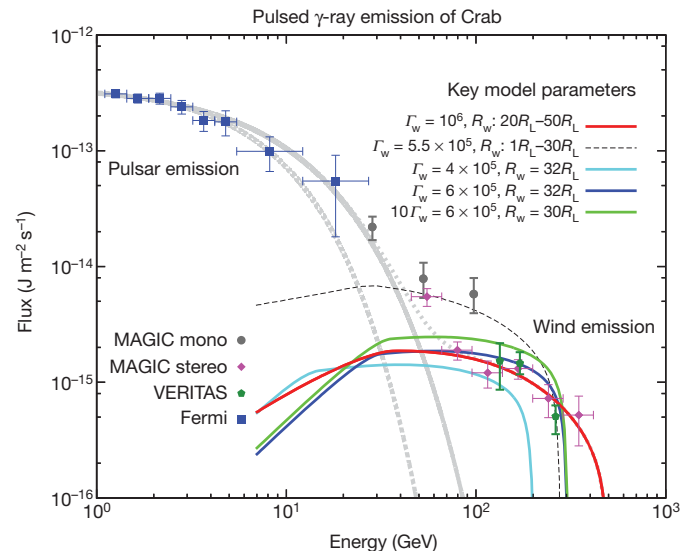


Figure 1 | Spectral energy distribution of γ -ray radiation produced by the pulsar magnetosphere and by the pulsar wind. Symbols show the reported γ -ray fluxes with 1-s.d. error bars^{7–10}. Curves show theoretical predictions (this work). The Fermi Large Area Telescope points¹⁰ are best fitted by the function $F_E = 3.8 \times 10^{-13} E^{0.03} \exp[-E/5.8 \text{ GeV}] \text{ J m}^{-2} \text{ s}^{-1}$ (dashed grey line). Assuming a slightly harder spectrum in the cut-off region, with $F_E = 3.8 \times 10^{-13} E^{0.03} \exp[-(E/7 \text{ GeV})^{0.85}] \text{ J m}^{-2} \text{ s}^{-1}$ (solid grey line), the MAGIC ‘mono’ data points⁸ can be explained as well (because of large systematic uncertainties, the mono 100-GeV point, which differs by a factor of three from the flux measured by two MAGIC telescopes in the more reliable stereoscopic regime⁹, perhaps ought to be discarded). This spectrum is somewhat harder than that predicted by standard magnetospheric models, but does not challenge them^{16–18}. The inverse-Compton γ -ray emission of the cold ultrarelativistic wind¹³ can naturally explain the pulsed γ -ray fluxes reported^{7,9} above 100 GeV. The solid light-blue, blue and green curves are calculated under the assumption of ‘instant’ acceleration of the wind at the fixed radius R_w . In principle, the acceleration can start earlier, but closer to the light cylinder the acceleration rate should be modest; otherwise it would lead to overproduction of inverse-Compton γ -rays. Earlier acceleration is demonstrated by the dashed black curve, which is calculated under the assumption that acceleration starts at the light cylinder with a rate that increases in proportion with R^3 up to $R_w = 30R_L$, where the Lorentz factor equals 5.5×10^5 (Supplementary Information). The solid red curve corresponds to the case in which the Poynting flux transformation takes place within the $20R_L\text{--}50R_L$ zone, assuming the wind’s acceleration rate to be independent of distance; the maximum Lorentz factor, achieved at $50R_L$, is set to 10^6 . (The dotted grey line corresponds to the superposition of the red and solid grey lines and shows the transition between the two radiation components.) Because of the decrease in the density of target photons with distance, the main fraction of VHE radiation is produced at around $30R_L$ with a Lorentz factor close to 5×10^5 . This explains the general similarity of the red curve to the instant-acceleration curves, apart from in the highest-energy region, where the sharp cut-off of the red curve is shifted to ~ 500 GeV.

¹Dublin Institute for Advanced Studies, School of Cosmic Physics, 31 Fitzwilliam Place, Dublin 2, Ireland. ²Max Planck Institute for Nuclear Physics, Saupfercheckweg 1, 69117 Heidelberg, Germany.

³National Research Nuclear University (MEPhI), Kashirskoe shosse 31, Moscow, 115409, Russia. ⁴Institute of Space and Astronautical Science/JAXA, 3-1-1 Yoshinodai, Chuo-ku, Sagami-hara, Kanagawa 252-5210, Japan.

reduction of the Compton cross-section due to the Klein–Nishina effect means that the X-ray band is the main contributor to the Comptonization of the wind. The X-ray flux is well measured up to 100 keV (ref. 14) and therefore the calculations of the inverse-Compton radiation depend basically on the site and the dynamics (speed) of transformation of the Poynting flux to kinetic energy of bulk motion.

We assume that at a distance R_w from the pulsar, the wind is accelerated to the Lorentz factor Γ_w (Fig. 2). Particles of the accelerated wind cannot move purely radially, because the wind should carry both the energy and the angular momentum lost by the pulsar. From the relation between the rotation energy (E_{rot}) and angular momentum (M_{rot}) losses, $\dot{E}_{\text{rot}} = \Omega \dot{M}_{\text{rot}}$, where Ω is the angular velocity of the rotating sphere and a dot denotes a time derivative, we can define the trajectory of the wind particles. Indeed, each particle of the wind carries energy $\Gamma_w mc^2$ and angular momentum $\Gamma_w m r_{\perp} v$, where m , r_{\perp} and v are the particle's mass, lever arm and speed, respectively, and c is the speed of light. Because $\Gamma_w m r_{\perp} v \Omega = \Gamma_w mc^2$, particles in the accelerated wind move along straight lines, tangent to the light cylinder. Therefore, all photons emitted by the magnetosphere will collide with electrons of the wind at a non-zero angle, θ , resulting in inverse-Compton γ -rays. The γ -ray production efficiency depends on the electron Lorentz factor, the density of the target photons and the interaction angle. Because the cold wind carries almost the entire spin-down luminosity, even a tiny efficiency of about $\kappa \approx 10^{-6}$ should be sufficient to produce detectable γ -rays at an energy flux level of $F_E = \kappa \dot{E}_{\text{rot}} / 4\pi d^2 \approx 10^{-15} \text{ J m}^{-2} \text{ s}^{-1}$, where $d \approx 6 \times 10^{19} \text{ m}$ is the distance to the Crab.

Generally, the light curve of the target photons should be reflected in the time structure of the inverse-Compton γ -ray signal; however, they cannot be identical, owing, for example, to the effects related to the specifics of the anisotropic inverse-Compton scattering. More importantly, the geometrical effects may lead to non-negligible differences between the arrival times of the target photon and the secondary γ -ray pulses (Fig. 3). For wind located close to the light cylinder, the γ -ray signal seems shifted in time relative to the reported γ -ray data, by $\Delta t \approx 0.1 T$. By contrast, for wind acceleration at $R_w = 30R_L$, the widths and the positions of the predicted and observed γ -ray peaks (P1 and P2, respectively) are in very good agreement. However, whereas in the case of the isotropic wind the predicted P1/P2 flux ratio of the γ -ray signal mimics the X-ray light curve¹⁵ (Fig. 3, black crosses), the reported γ -ray data^{7,9} seem to correspond to a smaller ratio, $P1/P2 < 1$. This can be explained by there being a non-negligible wind

anisotropy, which would introduce noticeable corrections to the shape of the γ -ray light curve in general and to the P1/P2 ratio in particular (Fig. 3). The large uncertainties in the present γ -ray data prevent us from reaching a strong conclusion in this regard, but the improvement of the quality of VHE γ -ray light curves should in future allow the strength and the character of the wind anisotropy to be decisively probed.

GeV γ -rays have a light curve¹⁰ that is essentially different from the reported VHE light curves^{7,9}. This can be interpreted as a result of the production of GeV and TeV γ -rays in regions well separated from each other. This conclusion is supported by the spectral energy distribution of the time-averaged GeV and TeV signals. As demonstrated in Fig. 1, the entire γ -ray region can be considered a superposition of two separate components. Indeed, by introducing a new, flat-spectrum VHE component of the Comptonized wind, in addition to the nominal (magnetospheric) GeV component, the reported data in the GeV-to-TeV energy intervals can be smoothly matched.

Although inverse-Compton γ -rays are produced by mono-energetic electrons, the spectral energy distribution of γ -rays in the range of tens to hundreds of GeV is quite flat. This is caused by the combination of effects related to the broad power-law distribution of seed photons and the transition of the Compton cross-section from the Thomson regime to the Klein–Nishina regime. On the other hand, the spectrum is expected to have a very sharp cut-off at $E = \Gamma_w mc^2$. This not only can serve as a distinct feature for the identification of the wind origin of γ -rays, but also should allow us to determine the Lorentz factor of the wind. In fact, the measurements available at present do not allow strong deviation of the Lorentz factor from 5×10^5 . We note that the calculations do not depend on the ‘magnetization parameter’ σ (the ratio of the electromagnetic energy flux to the kinetic energy flux) as long as $R_w \gg R_L$. However, formally we can explain the pulsed VHE emission even for $\sigma \geq 1$. In this case, the acceleration should occur closer to the pulsar ($R_w \propto 1/\sigma^{1/2}$) to compensate for the reduction in the wind's kinetic energy. But in this case, the inverse-Compton γ -ray radiation is expected to have quite different spectral and temporal features.

The above estimates of the location of wind's acceleration site and its Lorentz factor are quite robust, but they are obtained under the assumption that the transformation of the Poynting flux proceeds very quickly, at a specific radius between R_w and $R_w \pm \delta R_w$ with $\delta R_w/R_w \leq 1$. This is not an obvious assumption, but is instead a working hypothesis that the wind acceleration takes place in a narrow zone at the radius $R_w \approx 30R_L$. We cannot a priori exclude the possibility

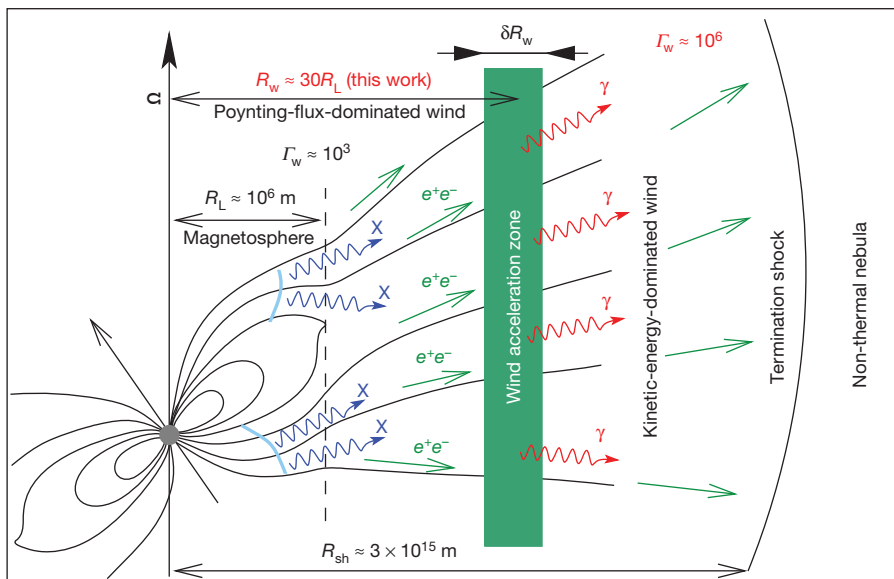


Figure 2 | Complex comprising the pulsar magnetosphere, the ultrarelativistic wind and the pulsar wind nebula. Dense electron (e^-)–positron (e^+) plasma produced in the pulsar magnetosphere by pair creation processes¹⁹ initiates an electron–positron wind at the light cylinder, which has radius $R_L \approx 10^6 \text{ m}$. Initially, the rotational energy lost by the pulsar, $\dot{E}_{\text{rot}} = 5 \times 10^{31} \text{ J s}^{-1}$, is released mainly in the form of electromagnetic energy (Poynting flux) and the wind's Lorentz factor therefore cannot be very large. At a distance R_w , the Poynting flux is converted to the kinetic energy of bulk motion (green zone), leading to an increase in the bulk-motion Lorentz factor to at least²⁰ $\Gamma_w \approx 10^4$. The termination of the wind by a standing reverse shock at $R_{\text{sh}} \approx 3 \times 10^{15} \text{ m}$ boosts the energy of the electrons to 10^{15} eV and randomizes their pitch angles². The radiative cooling of these electrons through the synchrotron and inverse-Compton processes results in an extended non-thermal source^{21–23}, the Crab nebula.

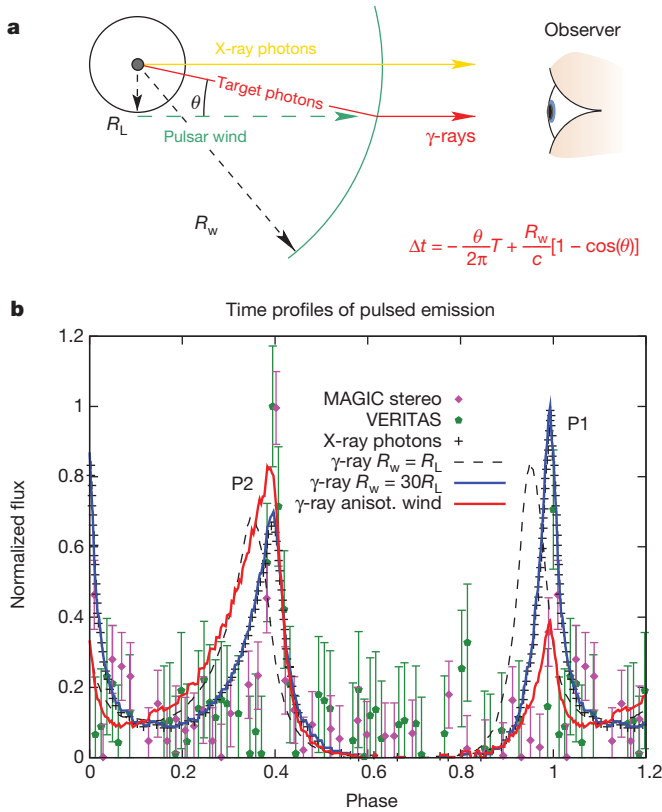


Figure 3 | Formation of the pulsed VHE inverse-Compton γ -ray signal in the wind of the Crab pulsar. **a**, Geometry of the inverse Compton scattering of magnetospheric X-rays by the electron-positron wind. **b**, Theoretical γ -ray light curves of the wind presented together with the reported VHE data^{7,9}. The velocity of the accelerated wind is tangential to the light cylinder (the direction of motion of electrons towards the observer is shown by the dashed green arrow). The interaction of electrons with the magnetospheric X-rays occurs predominantly at a distance $R \approx R_w$, where the wind is accelerated. Owing to the decrease in the target photon density with distance, the production of inverse-Compton γ -rays is suppressed at larger distances. The target X-ray photon converted to a VHE γ -ray photon reaches the observer earlier than an 'identical' photon emitted directly towards the observer. Two factors contribute to the time shift, Δt : the up-scattered X-ray photon is emitted by the pulsar earlier, by a time $\theta T/2\pi$, where T is the pulsar period; and it travels an additional path length of $R_w[1 - \cos(\theta)]$. For $R_w \gg R_L$, the time shift is negligibly small: $\Delta t \approx -(T/4\pi)R_L/R_w$. For acceleration of the isotropic pulsar wind at $R_w = 30R_L$, the γ -ray light curve (solid blue line) closely resembles the shape of the measured X-ray light curve¹⁵ (black crosses). For wind accelerated close to the light cylinder, the γ -ray light curve is shifted and somewhat broadened by comparison with wind accelerated at $R_w \gg R_L$. The anisotropy of the wind can also strongly deform the γ -ray light curve; in particular, it can change the ratio of the fluxes corresponding to peaks P1 and P2. The solid red line is calculated for an anisotropy factor proportional to the square of the sine of the angle between the line of sight and the direction of the magnetic momentum. This light curve seems to be in better agreement with the VERITAS⁷ and MAGIC⁹ points than the light curve corresponding to the fully isotropic wind, although the statistical and systematic uncertainties of observations (only Poisson error bars corresponding to the total count rates are shown on the plot) do not allow a definite conclusion in this regard.

that the wind is gradually accelerated starting from the edge of the magnetosphere, but our numerical calculations show that this cannot be the case (Fig. 1 and Supplementary Information). This is because the gradual acceleration would lead to a large number of high-energy electrons being accelerated close to the light cylinder and, consequently, to the prolific production of inverse-Compton γ -rays, in contradiction with the reported fluxes. Thus, the effective acceleration of the wind should start not much before the radius of $30R_L$ and not much beyond it. Such a case, assuming a linear acceleration rate of $I(R) = I_0 + a(R/R_L - 1)$ within the $20R_L$ – $50R_L$ radial interval and a

maximum Lorentz factor of 10^6 achieved at $50R_L$, is shown in Fig. 1. The corresponding γ -ray spectrum is smoother than the energy spectra predicted in the case of an instant acceleration, and better fits the VHE spectral points (Fig. 1) with the position of the sharp cut-off in the γ -ray spectrum shifted to 500 GeV. Although the wind acceleration within the $20R_L$ – $50R_L$ interval seems to be a physically more realistic scenario than an instant acceleration, this is still quite a narrow zone and the acceleration of the wind up to the Lorentz factor of 10^6 is therefore quite abrupt. This conclusion does not agree with those of alternative models, for example the so-called reconnection models of pulsar wind nebulae^{3,4} based on the assumption that the transformation of the Poynting flux to kinetic energy of bulk motion is a slow process that takes place over the entire region of the unshocked wind.

Received 5 September; accepted 15 December 2011.

Published online 15 February 2012.

- Rees, M. J. & Gunn, J. E. The origin of the magnetic field and relativistic particles in the Crab Nebula. *Mon. Not. R. Astron. Soc.* **167**, 1–12 (1974).
- Kennel, C. F. & Coroniti, F. V. Magnetohydrodynamic model of Crab nebula radiation. *Astrophys. J.* **283**, 710–730 (1984).
- Coroniti, F. V. Magnetically striped relativistic magnetohydrodynamic winds: the Crab Nebula revisited. *Astrophys. J.* **349**, 538–545 (1990).
- Lyubarsky, Y. & Kirk, J. G. Reconnection in the striped pulsar wind. *Astrophys. J.* **547**, 437–448 (2001).
- Vlahakis, N. Ideal magnetohydrodynamic solution to the σ problem in Crab-like pulsar winds and general asymptotic analysis of magnetized outflows. *Astrophys. J.* **600**, 324–337 (2004).
- Beskin, V. S. & Nokhrina, E. E. The effective acceleration of plasma outflow in the paraboloidal magnetic field. *Mon. Not. R. Astron. Soc.* **367**, 375–386 (2006).
- Aliu, E. *et al.* Detection of pulsed gamma rays above 100 GeV from the Crab pulsar. *Science* **334**, 69–72 (2011).
- Aleksić, J. *et al.* Observations of the Crab pulsar between 25 and 100 GeV with the MAGIC I telescope. *Astrophys. J.* **742**, 43 (2011).
- Aleksić, J. *et al.* Phase-resolved energy spectra of the Crab Pulsar in the range of 50–400 GeV measured with the MAGIC Telescopes. *Astron. Astrophys.* (submitted); preprint at (<http://arxiv.org/abs/1109.6124>) (2011).
- Abdo, A. A. *et al.* Fermi Large Area Telescope observations of the Crab pulsar and nebula. *Astrophys. J.* **708**, 1254–1267 (2010).
- Tang, A. P. S., Takata, J., Jia, J. J. & Cheng, K. S. A revisit of the phase-resolved X-ray and γ -ray spectra of the Crab pulsar. *Astrophys. J.* **676**, 562–572 (2008).
- Lyutikov, M., Otte, N. & McCann, A. The very-high energy emission from pulsars: a case for inverse Compton scattering. *Astrophys. J.* (submitted); preprint at (<http://arxiv.org/abs/1108.3824>) (2011).
- Bogovalov, S. V. & Aharonian, F. A. Very-high-energy γ radiation associated with the unshocked wind of the Crab pulsar. *Mon. Not. R. Astron. Soc.* **313**, 504–514 (2000).
- Kuiper, L. *et al.* The Crab pulsar in the 0.75–30 MeV range as seen by the CGRO COMPTEL. A coherent high-energy picture from soft X-rays to high-energy γ -rays. *Astron. Astrophys.* **378**, 918–935 (2001).
- Rots, A. H. *et al.* Absolute timing of the Crab pulsar with the Rossi X-Ray Timing Explorer. *Astrophys. J.* **605**, L129–L132 (2004).
- Hirofani, K. Outer-gap versus slot-gap models for pulsar high-energy emissions: the case of the Crab pulsar. *Astrophys. J.* **688**, 1254–1267 (2010).
- Osmanov, Z. & Rieger, F. M. On particle acceleration and very high energy γ -ray emission in Crab-like pulsars. *Astron. Astrophys.* **502**, L25–L28 (2008).
- Chkheidze, N., Machabeli, G. & Osmanov, Z. On the very high energy spectrum of the Crab pulsar. *Astrophys. J.* **730**, 62 (2011).
- Sturrock, P. A. A model of pulsars. *Astrophys. J.* **164**, 529–556 (1971).
- Wilson, D. B. & Rees, M. J. Induced Compton scattering in pulsar winds. *Mon. Not. R. Astron. Soc.* **185**, 297–304 (1978).
- Gould, R. J. High-energy photons from the Compton-synchrotron process in the Crab nebula. *Phys. Rev. Lett.* **15**, 577–579 (1965).
- de Jager, O. C. & Harding, A. K. The expected high-energy to ultra-high-energy γ -ray spectrum of the Crab nebula. *Astrophys. J.* **283**, 710–730 (1992).
- Atayan, A. M. & Aharonian, F. A. On the mechanisms of γ radiation in the Crab nebula. *Mon. Not. R. Astron. Soc.* **278**, 161–172 (1992).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We would like to thank J. Cortina, E. de Ona Wilhemi, S. Klepser and N. Otte for information about high-energy γ -ray observations. We also appreciate discussions with F. Rieger, D. Jones and P. Gandhi.

Author Contributions F.A.A., S.V.B. and D.K. jointly contributed in comparable proportions to all aspects of the work, including the calculations and preparation of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to F.A.A. (felix.aharonian@dias.ie).

The same pocket in menin binds both MLL and JUND but has opposite effects on transcription

Jing Huang^{1,2}, Buddha Gurung^{3*}, Bingbing Wan^{1,2*}, Smita Matkar³, Natalia A. Veniaminova⁴, Ke Wan^{1,2}, Juanita L. Merchant^{4,5}, Xianxin Hua³ & Ming Lei^{1,2}

Menin is a tumour suppressor protein whose loss or inactivation causes multiple endocrine neoplasia 1 (MEN1), a hereditary autosomal dominant tumour syndrome that is characterized by tumorigenesis in multiple endocrine organs¹. Menin interacts with many proteins and is involved in a variety of cellular processes^{2–8}. Menin binds the JUN family transcription factor JUND and inhibits its transcriptional activity^{2,9}. Several *MEN1* missense mutations disrupt the menin–JUND interaction, suggesting a correlation between the tumour-suppressor function of menin and its suppression of JUND-activated transcription^{2,10}. Menin also interacts with mixed lineage leukaemia protein 1 (MLL1), a histone H3 lysine 4 methyltransferase, and functions as an oncogenic cofactor to upregulate gene transcription and promote MLL1-fusion-protein-induced leukaemogenesis^{5,7,11,12}. A recent report on the tethering of MLL1 to chromatin binding factor lens epithelium-derived growth factor (LEDGF) by menin indicates that menin is a molecular adaptor coordinating the functions of multiple proteins¹³. Despite its

importance, how menin interacts with many distinct partners and regulates their functions remains poorly understood. Here we present the crystal structures of human menin in its free form and in complexes with MLL1 or with JUND, or with an MLL1–LEDGF heterodimer. These structures show that menin contains a deep pocket that binds short peptides of MLL1 or JUND in the same manner, but that it can have opposite effects on transcription. The menin–JUND interaction blocks JUN N-terminal kinase (JNK)-mediated JUND phosphorylation and suppresses JUND-induced transcription. In contrast, menin promotes gene transcription by binding the transcription activator MLL1 through the peptide pocket while still interacting with the chromatin-anchoring protein LEDGF at a distinct surface formed by both menin and MLL1.

The amino-terminal region of MLL1 interacts with menin^{12,14,15}. Isothermal titration calorimetry measurements showed that the menin-binding motif (residues 6–25) of MLL1 (MLL1_{MBM}) is necessary and sufficient for menin binding (Fig. 1a and Supplementary Fig. 1a–c).

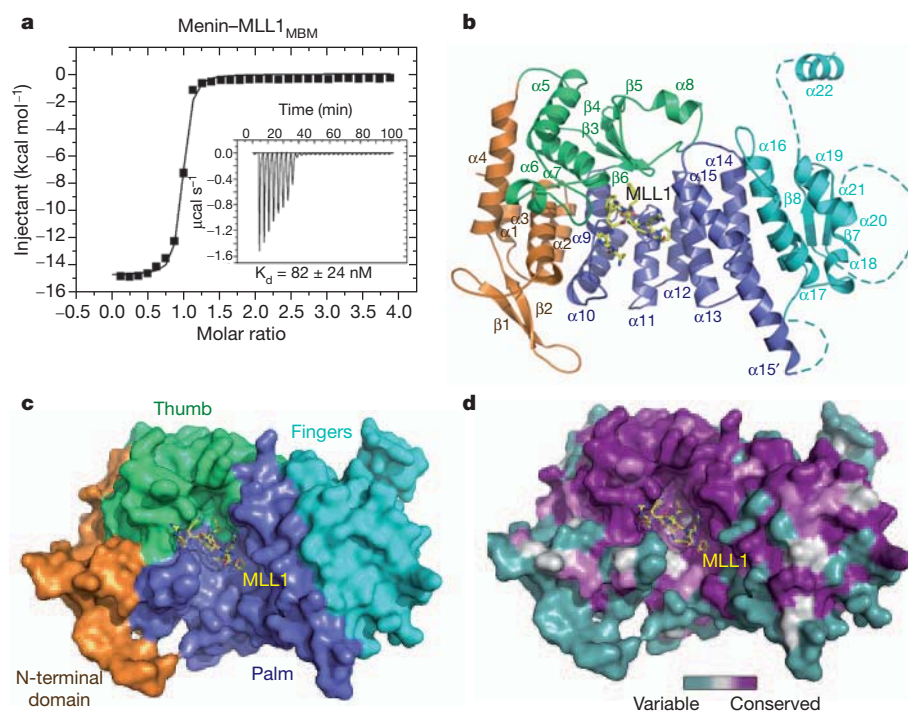


Figure 1 | Overview of the human menin–MLL1_{MBM} complex structure.

a, Isothermal titration calorimetry measurement of the menin–MLL1_{MBM} interaction. The inset shows the isothermal titration data. **b**, Overall structure of the menin–MLL1_{MBM} complex. The N-terminal domain is shown in orange, the thumb domain in green, the palm domain in blue, the fingers domain in cyan,

and loop regions that are disordered or not included in the crystal structure are shown as dashed lines. MLL1_{MBM} is shown as a stick model in yellow. **c**, The surface representation of menin indicates that menin adopts a curved left-hand-shaped conformation. **d**, Front view of the menin–MLL1_{MBM} complex, coloured according to the degree of amino acid conservation among menin homologues.

¹Howard Hughes Medical Institute, University of Michigan Medical School, 1150 West Medical Center Drive, Ann Arbor, Michigan 48109, USA. ²Department of Biological Chemistry, University of Michigan Medical School, 1150 West Medical Center Drive, Ann Arbor, Michigan 48109, USA. ³Abramson Family Cancer Research Institute, Department of Cancer Biology, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania 19104, USA. ⁴Department of Internal Medicine, University of Michigan, 109 Zina Pitcher Place, Ann Arbor, Michigan 48109, USA. ⁵Department of Molecular and Integrative Physiology, Division of Gastroenterology, University of Michigan, 109 Zina Pitcher Place, Ann Arbor, Michigan 48109, USA.

*These authors contributed equally to this work.

MLL2, the closest relative of MLL1, contains a sequence that is almost identical to MLL1_{MBM} at its N terminus (Supplementary Fig. 1b); MLL2_{16–35} (MLL2_{MBM}) binds to menin with an affinity that is comparable to that of MLL1_{MBM} (Supplementary Fig. 1d). To understand how MLL1 and MLL2 (collectively referred to as MLL) are recognized by menin, we determined the crystal structures of human menin alone

or in complex with MLL1_{MBM} (Supplementary Fig. 2, Supplementary Table 1 and Supplementary Information). The structure of human menin closely resembles a recently published menin homologue structure from *Nematostella*¹⁶.

The conformation of menin resembles a curved left 'hand' with a deep pocket formed by its 'thumb' and 'palm' (Fig. 1b, c). Menin consists

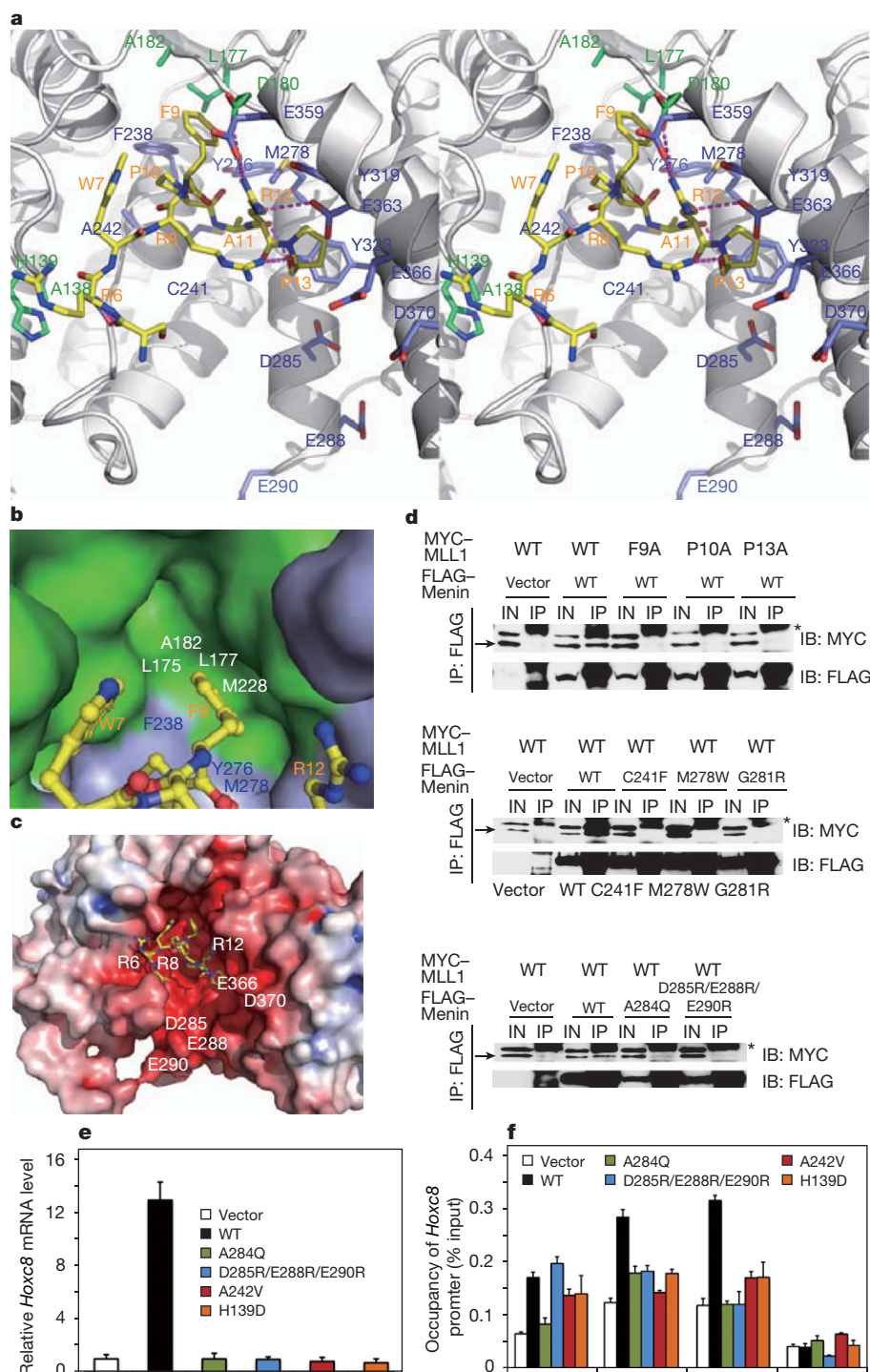


Figure 2 | Structural and mutational analyses of the menin-MLL1_{MBM} interaction. **a**, Stereo view of the menin-MLL1_{MBM} interface. The intermolecular hydrogen bonds are shown as dashed magenta lines. **b**, Phe⁹_{MLL1} (yellow) is nested in a hydrophobic pocket of menin formed by the thumb (green) and the palm (blue). **c**, Electrostatic surface potential of the MLL1_{MBM}-binding cavity of menin (positive potential, blue; negative potential, red). **d**, Co-immunoprecipitation of wild-type (WT) or mutant menin and

MLL1 proteins from 293T cells. Arrows and asterisks indicate the positions of MYC-MLL1 and immunoglobulin G (IgG), respectively. IB, immunoblot; IN, input; IP, immunoprecipitation. **e**, **f**, Expression of *Hoxc8* (**e**) and distributions of menin, MLL1 and H3K4me3 at the *Hoxc8* promoter (**f**) in *Men1*^{-/-} mouse embryonic fibroblasts (MEFs) complemented with control vector, WT or mutant menin (*n* = 6; error bar, standard deviation).

of four associated domains: an N-terminal domain characterized by a long β -hairpin, a transglutaminase-like domain that forms the thumb, a helical palm domain that contains three TPR motifs¹⁷ and a carboxy-terminal fingers domain (Fig. 1b, c, Supplementary Fig. 3 and Supplementary Information). Menin is highly conserved across species, and the conserved residues are either buried in the hydrophobic core or clustered together on a surface patch that covers the thumb and palm (Fig. 1d). MEN1 disease-derived missense and in-frame deletion mutations are evenly distributed throughout the protein (Supplementary Fig. 4), indicating that all four domains are important for the *in vivo* function of menin (Supplementary Fig. 4 and Supplementary Table 2).

The MLL1_{MBM} peptide adopts a compact conformation and plugs into the deep pocket of menin (Fig. 2a and Supplementary Fig. 5). Mutagenesis data indicates that MLL1_{MBM} residues Arg 6-Trp 7-Arg 8-Phe 9-Pro 10-Ala 11-Arg 12-Pro 13 and their interacting residues in menin contribute the most towards the interaction (Supplementary Figs 6 and 7, and Supplementary Tables 3 and 4). The side-chain of Phe 9^{MLL1} fits into a hydrophobic cavity formed by the thumb and palm of menin (Fig. 2b). A menin Met278Trp substitution altered the cavity shape and led to complete loss of binding (Supplementary Table 4). The MLL1_{MBM}-binding pocket is highly acidic (Fig. 2c). The two C-terminal arginine residues (Arg 24 and Arg 25) in MLL1_{MBM} are disordered, but they seem to be important for interaction, given that glutamate substitution resulted in a 21-fold decrease in binding affinity (Supplementary Table 3). Consistent with this, mutation of the acidic residues of menin also led to decreased binding (Supplementary Table 4).

Next, we examined the MLL1_{MBM}-binding activity of several MEN1 disease-derived mutations (His139Asp, Cys241Phe, Ala242Val, Gly281Arg, Ala284Gln, and Thr344Arg). Except for Ala284Gln and Thr344Arg, which yielded insoluble proteins, the remaining mutants impaired the menin-MLL1_{MBM} interaction (Supplementary Table 4). To further examine the menin-MLL1 interaction *in vivo*, we studied the interactions of mutant proteins that are transiently expressed in human embryonic kidney 293T cells. Consistent with the isothermal titration calorimetry analysis, co-immunoprecipitation data showed that mutations of the key residues at the interface completely abolished the menin-MLL1 interaction in cells (Fig. 2d).

Menin upregulates the expression of homeobox genes *Hoxc8* and *Hoxc6* (ref. 5). To test the effect of the menin-MLL interaction on the expression levels of *Hoxc8* and *Hoxc6*, wild-type and MLL-binding deficient mutants of menin were individually used to complement menin-null mouse embryonic fibroblasts. Western blot analyses indicated comparable expression of wild-type and mutant proteins in cells (Supplementary Fig. 8a). When *Men1*^{-/-} cells were complemented with wild-type menin, expression of *Hoxc8* and *Hoxc6* dramatically increased compared to vector-expressing cells (Fig. 2e and Supplementary Fig. 8b). In contrast, overexpression of the menin mutants in *Men1*^{-/-} cells failed to upregulate the messenger RNA levels of *Hoxc8* or *Hoxc6* (Fig. 2e and Supplementary Fig. 8b), suggesting that the menin-MLL interaction is essential for *Hoxc8* and *Hoxc6* expression.

Next we performed chromatin immunoprecipitation (ChIP) assays to determine the binding of mutant menin at the *Hoxc8* promoter. Except for Ala284Gln (a mutant that leads to insoluble proteins), all other mutants bound to the *Hoxc8* promoter as effectively as wild-type menin (Fig. 2f). Expression of wild-type or mutant menin did not greatly affect H3 distribution at the *Hoxc8* promoter (Supplementary Fig. 8c). Notably, *Men1*^{-/-} cells complemented with wild-type menin exhibited a substantial increase in MLL1 binding and histone H3K4me3 trimethylation at the *Hoxc8* promoter compared with vector-expressing or mutant-menin-expressing cells (Fig. 2f). Therefore, although menin mutants were able to bind to the *Hoxc8* promoter, their ability to recruit MLL1 and thus establish H3K4me3 at the *Hoxc8* promoter was compromised, resulting in reduced *Hoxc8* expression.

LEDGF, a chromatin-associated protein¹⁸, is required for MLL1-dependent transcription and leukaemic transformation¹³. Isothermal

titration calorimetry measurement showed that a complex composed of menin and an N-terminal fragment of MLL1, called MLL1_{MBM-LBM} (comprising residues 6–153 and including both menin-binding and LEDGF-binding motifs) binds to the integrase binding domain of LEDGF (LEDGF_{IBD}) with an affinity of 470 nM (Fig. 3a and Supplementary Fig. 9a). In contrast, neither menin nor MLL1_{MBM-LBM} alone could interact with LEDGF_{IBD} (Supplementary Fig. 9b)¹³. We determined the menin-MLL1_{MBM-LBM}-LEDGF_{IBD} complex structure at a resolution of 3.0 Å (Supplementary Fig. 9c and Supplementary Table 1). MLL1_{MBM-LBM} exhibits an extended conformation and binds to menin through two major sites (Fig. 3b); the N-terminal MLL1_{MBM} coil folds into the high-affinity pocket of menin in the same manner as in the menin-MLL1_{MBM} structure (Supplementary Figs 9d, e), whereas the C-terminal helix α 2 packs on the surface of the N-terminal domain of menin to form a V-shaped groove for LEDGF_{IBD} binding (Fig. 3b and Supplementary Fig. 9f). The middle loop of MLL1_{MBM-LBM} spans a large distance on menin without many specific interactions except for two leucine residues (Leu 106 and Leu 116) with side-chains that point to

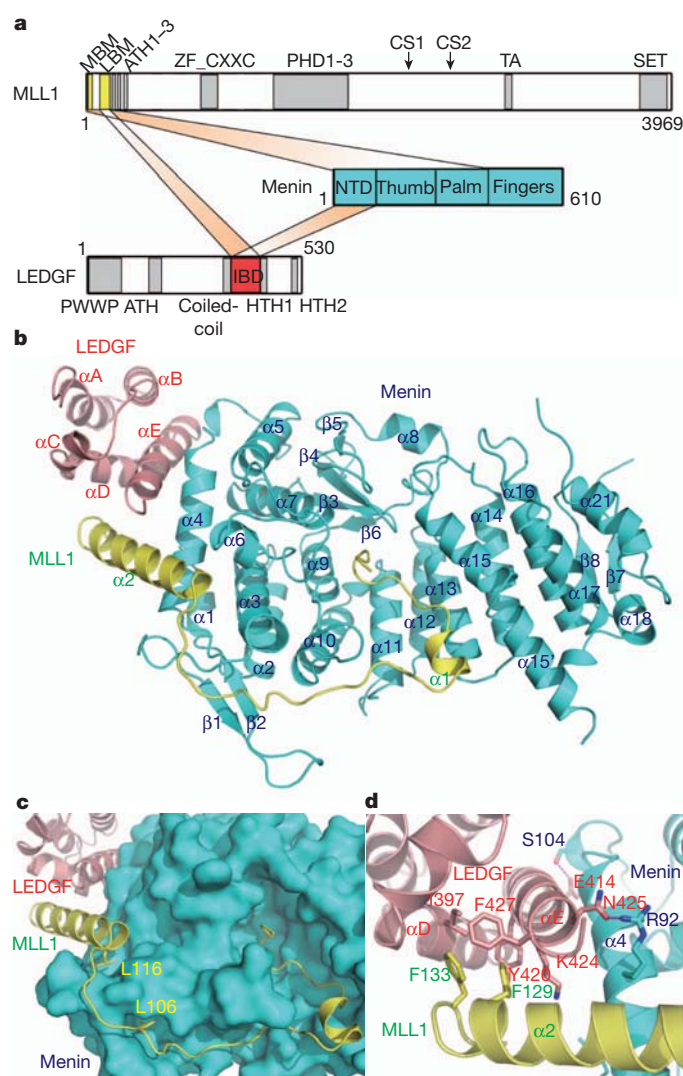


Figure 3 | Structure of the menin-MLL1_{MBM-LBM}-LEDGF_{IBD} ternary complex. **a**, Domain organization of menin, MLL1 and LEDGF. The menin- and LEDGF-binding motifs and LBM motifs of MLL1 are shown in yellow, menin in cyan, the integrase-binding domain of LEDGF in red and other regions in grey. Interactions among the three proteins are shown in orange. **b**, Ribbon diagram of the menin-MLL1_{MBM-LBM}-LEDGF_{IBD} complex. Menin is in cyan, MLL1_{MBM-LBM} in yellow and LEDGF_{IBD} in red. **c**, The extended MLL1 loop between MLL1_{MBM} and MLL1_{LBM} covers a large part of the surface area of menin. **d**, Detailed view of the intermolecular three-helix-bundle at the ternary interface.

two shallow pockets on the menin surface, defining the path of the loop (Fig. 3c and Supplementary Fig. 9g). Helix α E of LEDGF_{IBD} is sandwiched between helices α 2 of MLL1 and α 4 of menin through both hydrophobic and electrostatic interactions (Fig. 3d). In support of the crystal structure, mutations of residues on the α 4 helix of the N-terminal domain of menin (Ala95Arg and Ser104Tyr) specifically disrupted the interaction with LEDGF (Supplementary Table 5 and Supplementary Fig. 10). Notably, *Men1*^{-/-} cells that were complemented with these two mutants failed to stimulate *Hoxc8* expression (Supplementary Fig. 11), suggesting that a functional menin–MLL1–LEDGF complex is required for upregulation of *Hoxc8* expression. Together, our data show that menin functions as an adaptor molecule to modulate gene expression by binding MLL1 at one site while also interacting with LEDGF at a distinct surface.

Although MLL1 and MLL2 share many functional motifs, including the menin-binding motif (Supplementary Fig. 12), MLL2 does not contain a LEDGF-binding motif sequence and thus would not form a ternary complex with menin and LEDGF. Given that the PWWP domain of LEDGF, which contains a relatively well conserved Pro-Trp-Trp-Pro signature, is required for MLL1-mediated leukaemic transformation^{13,18}, the inability of MLL2 to form a menin–MLL2–LEDGF complex explains why only MLL1, and not MLL2, has so far been described as a proto-oncogene that can be activated by chromosomal translocations.

Menin also interacts directly with transcription factor JUND^{2,9}. We defined JUND residues 27–47 as the menin-binding motif (JUND_{MBM}) with an affinity of 1.6 μ M (Fig. 4a and Supplementary Fig. 13). Sequence comparison of JUND_{MBM} and MLL1_{MBM} revealed a

striking similarity (Fig. 4a), suggesting that JUND_{MBM} might interact with menin through the same binding pocket as does MLL1_{MBM}. Consistent with this idea, both isothermal titration calorimetry and glutathione S-transferase (GST) pull-down assays showed that MLL1 could efficiently compete with JUND for menin binding (Supplementary Fig. 14).

We determined the menin–JUND_{MBM} complex structure, which shows many similarities to the menin–MLL1_{MBM} structure (Fig. 4b and Supplementary Table 1). First, the Phe-Pro-(Ala or Gly)-(Arg or Ala)-Pro motifs in both menin-binding motifs are almost identical in overall conformation (Supplementary Fig. 15a). Second, Phe32, Pro33 and Pro36 of JUND interact with menin in the same way as their counterparts in MLL1_{MBM} (Supplementary Fig. 15b, c). Notably, two lysine residues (Lys 46 and Lys 47) in JUND_{MBM}, equivalent to the disordered Arg 24 and Arg 25 in MLL1_{MBM}, are visible in the electron density map and point to an acidic surface on menin (Supplementary Fig. 15c). Mutation of these lysine residues and other key binding residues at the interface abolished or weakened the interaction both *in vitro* and *in vivo* (Fig. 4c, Supplementary Fig. 16 and Supplementary Table 6).

Menin uncouples JUND phosphorylation from JNK activation, but the mechanism is poorly understood¹⁹. The consensus JNK-docking domain (D-domain) contains a cluster of basic amino acids preceding two leucine residues²⁰ (Fig. 4d). JUND_{MBM} is partially overlapped with a putative D-domain of JUND (JUND_D)²¹ (Fig. 4d). Both the basic residues and the leucine residues in JUND_D are indispensable for JNK docking on JUND as well as JNK-mediated JUND phosphorylation (Fig. 4e, f and Supplementary Fig. 17a, b). Thus, Lys 46 and Lys 47 are both required for menin binding and JUND phosphorylation by

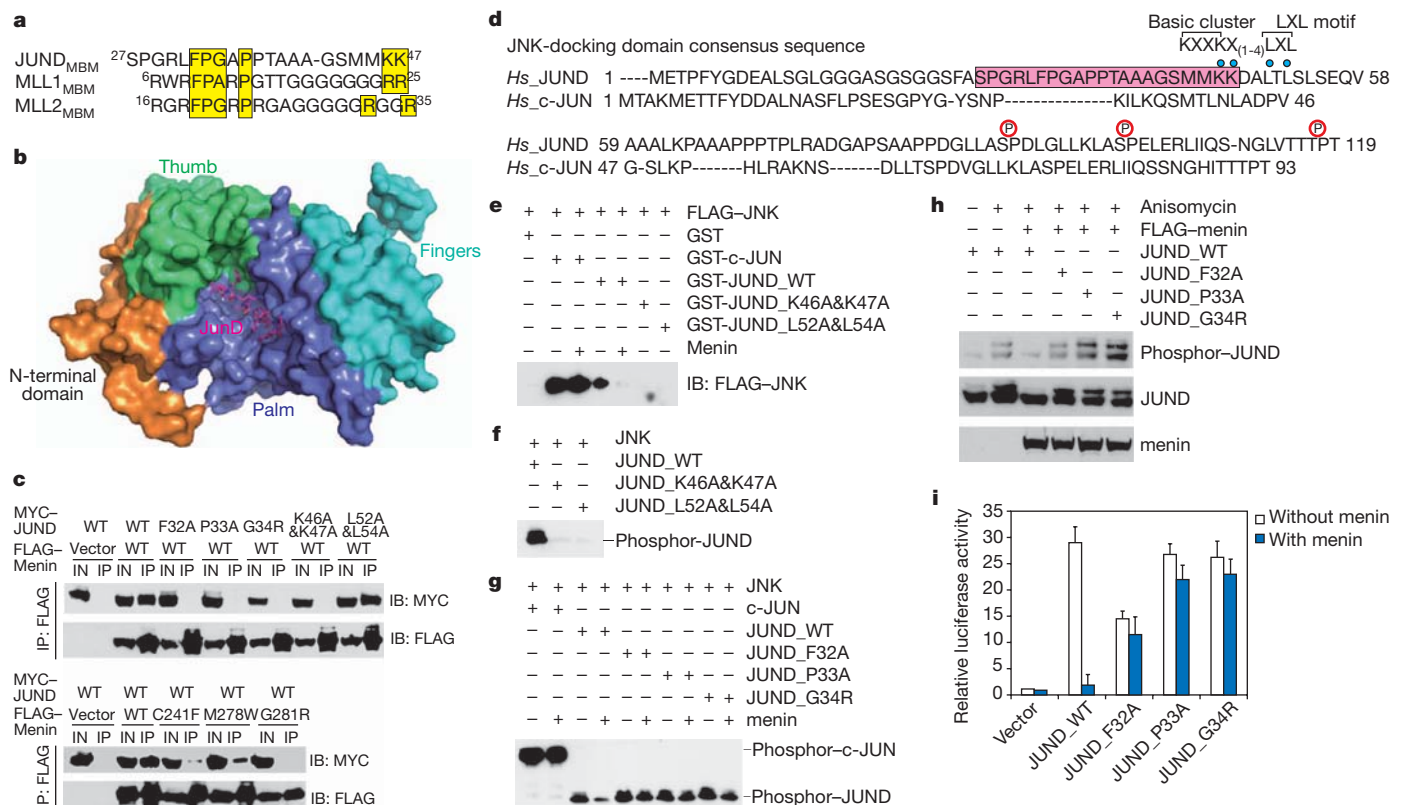


Figure 4 | Structural and functional studies of the menin–JUND interaction. **a**, Sequence alignment of the menin-binding motif sequences of JUND, MLL1 and MLL2. Conserved residues are highlighted in yellow. **b**, Crystal structure of the menin–JUND_{MBM} complex. Menin is coloured as in Fig. 1c and JUND_{MBM} is shown as a purple stick model. **c**, Co-immunoprecipitation of WT or mutant menin and JUND from 293T cells. **d**, Sequence comparison of the N termini of JUND and c-JUN. The menin-binding motif sequence of JUND is highlighted in purple. Key residues in the JNK-docking domain are denoted with blue dots and three phosphorylation

sites are labelled. **e**, *In vitro* GST-pull-down analysis of the interactions between FLAG-tagged JNK and the indicated JUND proteins. **f**, *In vitro* phosphorylation of WT or mutant JUND by JNK. **g**, *In vitro* phosphorylation of WT or menin-binding-deficient JUND by JNK in the presence or absence of menin. **h**, Menin suppresses JUND phosphorylation in response to anisomycin activation of JNK in 293T cells. **i**, WT or mutant JUND plasmids were transfected into 293T cells with AP1 and Renilla reporter plasmids, and with or without menin. Luciferase assays were performed 2 days after transfection ($n = 4$; error bar, standard deviation).

JNK. This led us to test whether menin inhibits JUND phosphorylation through sequestering JUND from JNK. In GST-pull-down assay, GST-JUND can only pull down JNK in the absence of menin, indicating that menin has a higher affinity to JUND (Fig. 4e and Supplementary Fig. 17a). Furthermore, when menin was added, phosphorylation of JUND was clearly inhibited (Fig. 4g and Supplementary Fig. 17c). In contrast, phosphorylation of menin-binding-deficient mutants of JUND was not affected by menin (Fig. 4g and Supplementary Fig. 17c). Next, we examined whether menin could inhibit JUND phosphorylation in response to anisomycin activation of JNK in 293T cells¹⁹. Although wild-type JUND phosphorylation was suppressed by menin, menin-binding deficient mutants remained robustly phosphorylated in the presence of menin (Fig. 4h). Notably, menin had no effects on JNK binding and JNK-mediated phosphorylation of c-JUN, a close homologue of JUND that lacks a menin-binding motif (Fig. 4d, e, g and Supplementary Fig. 17a, c, d). Together, our findings reveal that the menin-JUND interaction blocks JNK docking on JUND and inhibits the JNK-mediated phosphorylation.

Menin represses JUND-mediated transcriptional activation^{2,9}. To examine whether this repression depends on the menin-JUND interaction, wild-type or menin-binding-deficient mutants of JUND were co-transfected into 293T cells with an AP1 luciferase reporter plasmid in the presence or absence of menin (Supplementary Fig. 17e). Consistent with previous studies, transactivation by JUND was effectively repressed by menin^{2,9} (Fig. 4i). In contrast, menin exhibited a marginal effect on mutant JUND-mediated transcriptional activation (Fig. 4i). We recently demonstrated that JUND induces gastrin gene expression in human AGS gastric cells and that this induction can be suppressed by menin²². Consistent with the luciferase assay, menin failed to suppress the gastrin upregulation that was induced by mutant JUND, suggesting that the menin-JUND interaction is important in gastrin expression regulation (Supplementary Fig. 18). Thus, we conclude that the menin-JUND interaction plays a key part in suppressing JUND-mediated transcriptional activation.

In summary, our structural and functional studies provide a mechanistic explanation of how menin could both positively and negatively regulate gene transcription. Our findings also provide evidence that menin acts as a scaffold protein to assemble a menin-MLL1-LEDGF ternary complex to coordinate gene transcription and promote MLL1-fusion-protein-induced leukaemogenesis.

METHODS SUMMARY

Human menin, LEDGF_{IBD} and the MLL and JUND peptides were expressed in *Escherichia coli* BL21(DE3) and purified by sequential affinity and gel-filtration chromatography purification. Menin crystals were obtained in sitting drops over 100 mM sodium cacodylate (pH 6.5) and 1.4 M sodium acetate. Crystallization of menin with the MLL_{1MBM} or JUND_{MBM} peptides was achieved by sitting-drop diffusion with a well solution containing 100 mM Tris-HCl (pH 7.0), 200 mM MgCl₂ and 2.3 M NaCl. The menin-MLL_{1MBM}-LBM-LEDGF_{IBD} complex was crystallized by hanging-drop vapour diffusion against a well solution of 50 mM HEPES (pH 7.0), 1.6 M (NH₄)₂SO₄, 10 mM MgCl₂, 0.016% L-canavanine, 0.016% O-phospho-L-serine, 0.016% taurine, 0.016% quinine, 0.016% sodium glyoxylate monohydrate and 0.016% cholic acid, and were dehydrated with the solution containing 50 mM HEPES (pH 7.0), 2.3 M (NH₄)₂SO₄ and 10 mM MgCl₂. The menin-MLL_{1MBM} complex structure was determined by multi-wavelength anomalous dispersion to a resolution of 3.0 Å. The structures of menin alone, the menin-MLL_{1MBM}-LBM-LEDGF_{IBD} ternary complex, and the menin-JUND_{MBM} complex were solved by molecular replacement and refined to resolutions of 2.5 Å, 3.0 Å and 2.85 Å, respectively.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 17 July; accepted 19 December 2011.

Published online 12 February 2012.

- Chandrasekharappa, S. C. *et al.* Positional cloning of the gene for multiple endocrine neoplasia-type 1. *Science* **276**, 404–407 (1997).

- Agarwal, S. K. *et al.* Menin interacts with the AP1 transcription factor JunD and represses JunD-activated transcription. *Cell* **96**, 143–152 (1999).
- Busygina, V., Kottemann, M. C., Scott, K. L., Plon, S. E. & Bale, A. E. Multiple endocrine neoplasia type 1 interacts with forkhead transcription factor CHES1 in DNA damage response. *Cancer Res.* **66**, 8397–8403 (2006).
- Chen, G. *et al.* Menin promotes the Wnt signaling pathway in pancreatic endocrine cells. *Mol. Cancer Res.* **6**, 1894–1907 (2008).
- Hughes, C. M. *et al.* Menin associates with a trithorax family histone methyltransferase complex and with the *Hoxc8* locus. *Mol. Cell* **13**, 587–597 (2004).
- Jin, S. *et al.* Menin associates with FANCD2, a protein involved in repair of DNA damage. *Cancer Res.* **63**, 4204–4210 (2003).
- Yokoyama, A. *et al.* Leukemia proto-oncoprotein MLL forms a SET1-like histone methyltransferase complex with menin to regulate *Hox* gene expression. *Mol. Cell Biol.* **24**, 5639–5649 (2004).
- Yang, Y. & Hua, X. In search of tumor suppressing functions of menin. *Mol. Cell. Endocrinol.* **265–266**, 34–41 (2007).
- Knapp, J. I. *et al.* Identification and characterization of JunD missense mutants that lack menin binding. *Oncogene* **19**, 4706–4712 (2000).
- Groussin, L. & Bertherat, J. Mechanisms of multiple endocrine neoplasia type 1: evidence for regulation of the AP-1 family of transcription factors by menin. *Eur. J. Endocrinol.* **141**, 15–16 (1999).
- Krivtsov, A. V. & Armstrong, S. A. MLL translocations, histone modifications and leukaemia stem-cell development. *Nature Rev. Cancer* **7**, 823–833 (2007).
- Yokoyama, A. *et al.* The menin tumor suppressor protein is an essential oncogenic cofactor for MLL-associated leukemogenesis. *Cell* **123**, 207–218 (2005).
- Yokoyama, A. & Cleary, M. L. Menin critically links MLL proteins with LEDGF on cancer-associated target genes. *Cancer Cell* **14**, 36–46 (2008).
- Caslini, C. *et al.* Interaction of MLL amino terminal sequences with menin is required for transformation. *Cancer Res.* **67**, 7275–7283 (2007).
- Grembecka, J., Belcher, A. M., Hartley, T. & Cierpicki, T. Molecular basis of the mixed lineage leukemia-menin interaction: implications for targeting mixed lineage leukemias. *J. Biol. Chem.* **285**, 40690–40698 (2010).
- Murai, M. J., Chruszcz, M., Reddy, G., Grembecka, J. & Cierpicki, T. Crystal structure of menin reveals binding site for mixed lineage leukemia (MLL) protein. *J. Biol. Chem.* **286**, 31742–31748 (2011).
- Lamb, J. R., Tugendreich, S. & Hieter, P. Tetratricopeptide repeat interactions: to TPR or not to TPR? *Trends Biochem. Sci.* **20**, 257–259 (1995).
- Llano, M., Morrison, J. & Poeschla, E. M. Virological and cellular roles of the transcriptional coactivator LEDGF/p75. *Curr. Top. Microbiol. Immunol.* **339**, 125–146 (2009).
- Gallo, A. *et al.* Menin uncouples Elk-1, JunD and c-Jun phosphorylation from MAP kinase activation. *Oncogene* **21**, 6434–6445 (2002).
- Yang, S. H., Whitmarsh, A. J., Davis, R. J. & Sharrocks, A. D. Differential targeting of MAP kinases to the ETS-domain transcription factor Elk-1. *EMBO J.* **17**, 1740–1749 (1998).
- Yazgan, O. & Pfarr, C. M. Regulation of two JunD isoforms by Jun N-terminal kinases. *J. Biol. Chem.* **277**, 29710–29718 (2002).
- Mensah-Osman, E. J., Veniaminova, N. A. & Merchant, J. L. Menin and JunD regulate gastrin gene expression through proximal DNA elements. *Am. J. Physiol. Gastrointest. Liver Physiol.* **301**, G783–G790 (2011).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank G. Wilding for the human JUND complementary DNA and P. Cherepanov for the human LEDGF cDNA. We thank Y. Chen and W. Deng for help at various stages of the project. M.L. is a Howard Hughes Medical Institute Early Career Scientist. Work was supported by National Institutes of Health grants (GM 083015-01 to M.L., R01-DK085121 to X.H. and R37-DK45729 to J.L.M.), an American Cancer Society Research Scholar grant (to M.L.) and an American Association for Cancer Research Caring for Carcinoid Foundation Grant (to X.H.). The General Medicine and Cancer Institutes Collaborative Access Team has been funded in whole or in part with federal funds from the National Cancer Institute (grant Y1-CO-1020) and the National Institute of General Medical Science (grant Y1-GM-1104). Use of the Advanced Photon Source was supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences, under contract number DE-AC02-06CH11357.

Author Contributions J.H. is responsible for the bulk of the experiments, B.G. for the qRT-PCR and ChIP assays; B.W. for the co-immunoprecipitation and *in vivo* phosphorylation analyses, S.M. for the luciferase assay, N.A.V. and J.L.M. for the analysis of gastrin expression, and K.W. for some of the protein purification. M.L. and X.H. supervised the project and wrote the paper.

Author Information The atomic coordinates and structure factors of menin, menin-MLL_{1MBM}, menin-JUND_{MBM}, and menin-MLL_{1MBM}-LBM-LEDGF_{IBD} have been deposited in the RCSB Protein Data Bank under accession codes 3U84, 3U85, 3U86 and 3U88, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.L. (leim@umich.edu) or X.H. (huax@mail.med.upenn.edu).

METHODS

Protein expression and purification. To facilitate crystallization, we genetically deleted an unstructured loop (residues 460–519) in menin, a short fragment (residues 40–45) in JUND_{MBM} and two loop regions (residues 16–22 and 36–102) in MLL1_{MBM-LBM}. All the resulting proteins retain wild-type-like binding affinities (Supplementary Figs 2d, 9c and 13d). For simplicity, MeninΔ, JUND_{MBM}Δ and MLL1_{MBM-LBM}Δ are referred to as menin, JUND_{MBM} and MLL1_{MBM-LBM}, respectively, unless stated otherwise.

Various human menin proteins and the MLL and JUND peptides were expressed in *E. coli* BL21(DE3) using a modified pET28b vector with a SUMO protein fused at the N terminus after the His₆ tag. After induction for 16 h with 0.1 mM isopropylthiogalactoside (IPTG) at 25 °C, the cells were collected by centrifugation and the pellets were resuspended in lysis buffer (50 mM Tris-HCl, pH 8.0; 50 mM NaH₂PO₄; 400 mM NaCl; 3 mM imidazole; 10% glycerol; 0.1 mg ml⁻¹ lysozyme; 2 mM 2-mercaptoethanol; 1 mM PMSF; 5 mM benzamide; 1 μg ml⁻¹ leupeptin; and 1 μg ml⁻¹ pepstatin). The cells were then lysed by sonication and the cell debris was removed by ultracentrifugation. The supernatant was mixed with Ni-NTA agarose beads (Qiagen) and rocked for 2 h at 4 °C before elution with 250 mM imidazole. Ulp1 protease was then added to remove the His₆-SUMO tag. After Ulp1 digestion, the menin proteins and the MLL and JUND peptides were further purified by gel-filtration chromatography on Hiload Superdex 200 and Hiload Superdex 75 columns (GE Healthcare), equilibrated with buffer A (25 mM Tris-HCl, pH 8.0; 150 mM NaCl; and 5 mM dithiothreitol (DTT)) and buffer B (100 mM ammonium bicarbonate), respectively. The purified menin proteins were concentrated to 25 mg ml⁻¹ and stored at -80 °C. The purified peptides were lyophilized and resuspended in water at a concentration of 50 mg ml⁻¹ and stored at -80 °C.

For the menin-MLL1_{MBM-LBM}-LEDGF_{IBD} complex, we cloned LEDGF_{IBD} into a modified pET28b vector with a SUMO protein fused at the N terminus after the His₆ tag. MLL1_{MBM-LBM} was cloned into a GST fusion protein expression vector, pGEX6p-1 (GE healthcare). The menin-MLL1_{MBM-LBM} complex and LEDGF_{IBD} itself were expressed in *E. coli* BL21(DE3), respectively. The menin-MLL1_{MBM-LBM} complex was purified by sequential affinity chromatography with Ni-NTA agarose beads and glutathione sepharose 4B beads (GE Healthcare). After removal of the His₆-UMO tag and GST tag with Ulp1 and Protease 3C, respectively, the complex was purified further with gel-filtration chromatography on a Hiload Superdex 200. Meanwhile, LEDGF_{IBD} was purified in the same way as menin and then mixed with the purified menin-MLL1_{MBM-LBM} complex with a molar ratio of 2:1. After 1 h incubation on ice, the protein mixtures were purified again with gel-filtration chromatography on a Hiload Superdex 200 column.

For the *in vitro* assays, mutant menin proteins were expressed in *E. coli* and purified following the procedure described above. All the mutant menin proteins displayed unaltered biophysical properties as analysed by gel-filtration chromatography (data not shown), ensuring that the altered affinities of the menin mutants for MLL1_{MBM}, MLL1_{MBM-LBM}-LEDGF_{IBD} and JUND_{MBM} are not attributable to a change in the structural integrity of the resulting proteins.

Crystallization, data collection and structure determination. Menin was crystallized by sitting-drop vapour diffusion at 4 °C. The precipitant solution contained 100 mM sodium cacodylate trihydrate (pH 6.5) and 1.4 M sodium acetate trihydrate. For the menin-MLL1_{MBM} complex, purified menin was first mixed with the MLL1_{MBM} peptide at a molar ratio of 1:2 and then the mixture was incubated on ice for 1 h to allow complex formation. Crystallization of the complex was achieved by sitting-drop vapour diffusion at 4 °C with the well solution containing 100 mM Tris-HCl (pH 7.0), 200 mM MgCl₂ and 2.3 M NaCl. A similar procedure was also used for crystallization of the menin-JUND_{MBM} complex. The menin-MLL1_{MBM-LBM}-LEDGF_{IBD} complex was crystallized by hanging-drop vapour diffusion at 4 °C with the well solution containing 50 mM HEPES (pH 7.0), 1.6 M (NH₄)₂SO₄, 10 mM MgCl₂, 0.016% L-canavanine, 0.016% O-phospho-L-serine, 0.016% taurine, 0.016% quinine, 0.016% sodium glyoxylate monohydrate and 0.016% cholic acid. The crystals were then dehydrated with the solution containing 0.05 M HEPES (pH 7.0), 2.3 M (NH₄)₂SO₄ and 0.01 M MgCl₂.

All of the crystals were gradually transferred into a harvesting solution containing the respective precipitant solutions plus 5 M sodium formate, before being flash-frozen in liquid nitrogen for storage. Data were collected under cryogenic conditions (100 K). Selenomethionine-multi-wavelength anomalous dispersion data set of the menin-MLL1_{MBM} complex at the Se peak and inflection wavelengths were collected at the Advanced Photon Source (APS) beamline 21-ID-D and processed using HKL2000 (ref. 23). Seven selenium atoms were located and refined, and the multiwavelength anomalous diffraction data phases were calculated using SHARP²⁴. The initial multi-wavelength anomalous dispersion map of the menin-MLL1_{MBM} complex was substantially improved by solvent flattening. A model was manually built into the modified experimental electron density using O (ref. 25) and further refined in Phenix²⁶. Native data sets of menin and the menin

complexes were collected at the APS beamline 21-ID-D and processed using HKL2000. The structures were determined by molecular replacement using Phaser in the CCP4i suite²⁷ and further refined in Phenix. The majority (~95%) of the residues in all structures lie in the most favoured region in the Ramachandran plot, and the remaining structures lie in the additionally stereochemically allowed regions in the Ramachandran plot.

Isothermal titration calorimetry. The equilibrium dissociation constants of the menin-MLL1_{MBM}, menin-JUND_{MBM} and menin-MLL1_{MBM-LBM}-LEDGF_{IBD} interactions were determined using a VP-ITC calorimeter (MicroCal). The binding enthalpies were measured at 20 °C in 25 mM Tris-HCl (pH 8.0) and 150 mM NaCl. Two independent experiments were performed for every interaction described here. Isothermal titration calorimetry data were subsequently analysed and fitted using Origin 7 software (OriginLab) with blank injections of peptides into buffer subtracted from the experimental titrations before data analysis.

Yeast two-hybrid assay. The yeast two-hybrid assays were performed using the yeast L40 strain harbouring pBTM116 and pACT2 (Clontech) fusion plasmids. The colonies containing both plasmids were selected on -Leu -Trp plates. The activities of β-galactosidase were measured according to Clontech MATCHMAKER library protocol and the averages from three individual transformants were reported.

Plasmid construction. To generate recombinant retroviruses, pMX-2× FLAG-menin was constructed by inserting polymerase chain reaction (PCR)-amplified menin cDNA into the *Bam*HI/*Not*I site of the retroviral vector pMX-2× FLAG. To generate menin mutants, pMX-2× FLAG -menin was used as a template for site-directed mutagenesis using the QuikChange kit from Agilent.

Cell culture and transfection. Menin-null MEFs, HEK293T and the human AGS gastric adenocarcinoma cell line were cultured in Dulbecco's modified Eagle's medium complemented with 10% fetal calf serum and 1% PenStrep. Menin-null MEFs were infected with empty vector, wild-type or mutant menin-expressing retroviruses and were subjected to puromycin selection (2 μg ml⁻¹) 72 h post-infection for 2 days. AGS and 293T cells were transiently transfected with the indicated expression vectors using Lipofectamine 2000 (Invitrogen) for 48 h.

Co-immunoprecipitation. Human 293T cells were transfected with pcDNA3.1 vectors encoding c-MYC-tagged MLL1 (residues 1–153) and FLAG-tagged menin. Two days after transfection the cells were resuspended in 1 ml of lysis buffer (20 mM Tris-HCl, pH 7.5; 150 mM NaCl; 1.0% Triton X-100; 1 mM EDTA; and protease inhibitor cocktail). Immunoprecipitation of lysates was conducted using 20 μl anti-FLAG M2 affinity agarose (Sigma). After washing with lysis buffer, immunoprecipitated proteins were eluted with ×2 loading buffer (50 mM Tris-HCl, pH 6.8; 2% SDS; 10% 2-mercaptoethanol; 10% glycerol; and 0.002% bromophenol blue), subjected to protein gel-electrophoresis using 4–20% SDS-polyacrylamide gel electrophoresis (SDS-PAGE) and then transferred to a polyvinylidene fluoride (PVDF) membrane. After blocking with TBST buffer containing 5% skimmed milk, proteins on the membrane were detected by western blot using anti-FLAG (Sigma) and anti-c-MYC (Santa Cruz Biotechnology) antibodies. The same procedure was also used for the co-immunoprecipitation experiments for menin and JUND.

Quantitative real-time PCR analysis. Exponentially growing MEFs were seeded at 2 × 10⁵ cells per 100-mm dish and harvested 2 days later. AGS cells were transfected with the menin and JUND expression vectors for 48 h. Total RNA was isolated with an RNeasy minikit from Qiagen. Quantitative real-time PCR (qRT-PCR) was performed in an ABI 7500 Real Time PCR system (Applied Biosystems).

ChIP assay. MEFs were cross-linked with 1% formaldehyde for 10 min at 37 °C. Cross-linking was stopped by addition of 125 mM glycine. The ChIP assay was performed using the QuikCHIP kit from Imgenex, according to the manufacturer's instructions. Antibodies used for ChIP were anti-menin (Bethyl labs), anti-MLL1, anti-histone H3K4me3, anti-histone H3 and IgG (Abcam). Antibody-precipitated DNA-protein complex was reverse cross-linked, and the DNA was isolated using phenol-chloroform extraction and the precipitated DNA was used as the template for PCR.

GST-pull-down assay. GST, GST-fused c-JUN (residues 1–246), GST-fused JUND (residues 1–150) and FLAG-tagged JNK3 were expressed in *E. coli* BL21(DE3) and were purified to homogeneity. GST-pull-down assays were performed by incubating 10 μg of GST or GST-JUN, 10 μg of FLAG-JNK3 with 10 μl of glutathione sepharose 4B beads and either with or without 20 μg of full-length menin in binding buffer (50 mM Tris-HCl (pH 8.0) and 150 mM NaCl) at 4 °C overnight. The beads were then extensively washed with binding buffer four times and the bound proteins were eluted with 10 mM reduced glutathione in binding buffer. After separation on 15% SDS-PAGE and Ponceau S staining, FLAG-tagged JNK3 protein was detected by western blot using anti-FLAG antibody.

In vitro kinase assay. WT c-JUN (residues 1–246) and WT or mutant JUND proteins (residues 1–150) were expressed in *E. coli* BL21(DE3) and purified as

described above for the purification of menin. For the *in vitro* kinase assay, 2 µg substrate was mixed with 0.5 µg kinase and 50 µM ATP, either with or without 10 µg of full-length menin protein in the kinase buffer (50 mM Tris-HCl, pH 7.5; 20 mM MgCl₂; 20 mM β-glycerophosphate; 2 mM DTT; and 0.1 mM sodium orthovanadate), and incubated at 30 °C for 1 h. The reaction mixtures were then separated on 15% SDS-PAGE, visualized with Ponceau S staining and the phosphorylated JUN proteins were detected with anti-JUND phosphor-Ser100 (anti-c-JUN phosphor-Ser 73) antibody (Cell Signaling).

In vivo kinase assay. 293T cells were transfected with expression vectors encoding FLAG-tagged menin and c-MYC tagged JUND. After 48 h of transfection, cells were incubated for 30 min with or without 10 µg ml⁻¹ anisomycin (Sigma), a potent JNK activator, and then the cell lysates were subjected to western blot with anti-JUND phosphor-Ser100 (anti-c-JUN phosphor-Ser 73, Cell Signaling), anti-FLAG and anti-c-MYC antibodies.

Luciferase assay. 293T cells were transfected with 1 µg of AP1 luciferase reporter plasmid (Stratagene), which contains seven copies of AP1-binding consensus 12-O-tetradecanoylphorbol 13-acetate-response element (TRE) upstream of the luciferase reporter gene, 0.25 µg of Renilla reporter plasmid and 0.5 µg of WT or mutant JUND plasmids, either without or with 0.5 µg of menin cDNA. Luciferase assays were performed using the dual luciferase assay kit (Promega) 2 days after transfection. To determine the protein expression in each transfection, 20 µg of cell lysates were immunoblotted with anti-menin (Bethyl Laboratories) and anti-JUND (Santa Cruz Biotechnology) antibodies.

Cell fractionation. 10⁸ 293T cells were collected and washed in cold PBS and hypotonic buffer (10 mM Tris-HCl, pH 7.3; 10 mM KCl; 1.5 mM MgCl₂; 0.2 mM PMSF; and 10 mM β-mercaptoethanol). The cells were then allowed to swell for 15 min in hypotonic buffer. The swelled cells were then homogenized with glass Dounce homogenizer (Wheaton) using the loose pestle until cell membrane lysis was 80–90%. The nuclei were collected by centrifuging for 15 min at 3,300g, resuspended in high salt buffer (600 mM KCl; 20 mM Tris pH 7.4; 25% glycerol; 1.5 mM MgCl₂; and 0.2 mM EDTA) and homogenized to break the nuclear membrane. The nuclear extracts were collected by centrifugation at 25,000g for 30 min and were then fractionated on a Superose 6 gel-filtration column (GE Healthcare). The resulting fractions were resolved by 10% SDS-PAGE and probed with anti-menin, anti-MLL1 and anti-JUND antibodies.

23. Otwinowski, Z. & Minor, W. in *Methods in Enzymology* Vol. 26 (eds Carter, C. W. Jr & Sweet, R. M.) 307–326 (Academic Press, 1997).
24. de La Fortelle, E. & Bricogne, G. Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* **276**, 472–494 (1997).
25. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
26. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
27. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).

Control of ground-state pluripotency by allelic regulation of *Nanog*

Yusuke Miyanari¹ & Maria-Elena Torres-Padilla¹

Pluripotency is established through genome-wide reprogramming during mammalian pre-implantation development, resulting in the formation of the naive epiblast. Reprogramming involves both the resetting of epigenetic marks and the activation of pluripotent-cell-specific genes such as *Nanog* and *Oct4* (also known as *Pou5f1*)^{1–4}. The tight regulation of these genes is crucial for reprogramming, but the mechanisms that regulate their expression *in vivo* have not been uncovered. Here we show that *Nanog*—but not *Oct4*—is monoallelically expressed in early pre-implantation embryos. *Nanog* then undergoes a progressive switch to biallelic expression during the transition towards ground-state pluripotency in the naive epiblast of the late blastocyst. Embryonic stem (ES) cells grown in leukaemia inhibitory factor (LIF) and serum express *Nanog* mainly monoallelically and show asynchronous replication of the *Nanog* locus, a feature of monoallelically expressed genes⁵, but ES cells activate both alleles when cultured under 2i conditions, which mimic the pluripotent ground state *in vitro*. Live-cell imaging with reporter ES cells confirmed the allelic expression of *Nanog* and revealed allelic switching. The allelic expression of *Nanog* is regulated through the fibroblast growth factor–extracellular signal-regulated kinase signalling pathway, and it is accompanied by chromatin changes at the proximal promoter but occurs independently of DNA methylation. *Nanog*-heterozygous blastocysts have fewer inner-cell-mass derivatives and delayed primitive endoderm formation, indicating a role for the biallelic expression of *Nanog* in the timely maturation of the inner cell mass into a fully reprogrammed pluripotent epiblast. We suggest that the tight regulation of *Nanog* dose at the chromosome level is necessary for the acquisition of ground-state pluripotency during development. Our data highlight an unexpected role for allelic expression in controlling the dose of pluripotency factors *in vivo*, adding an extra level to the regulation of reprogramming.

The development of the naive epiblast in the late blastocyst is orchestrated by pluripotency-associated transcription factors such as NANOG and OCT4. OCT4 is required for establishing and maintaining the pluripotent state³, and NANOG is essential for the acquisition of pluripotency^{1,6}. In ES cells, the pluripotent state is also governed by these core transcription factors. *Nanog* occupies a central position in this network, but the mechanisms regulating its expression are unclear.

To address how pluripotency-associated factors are regulated during reprogramming *in vivo*, we assayed the *de novo* gene expression of *Nanog* and *Oct4* in mouse pre-implantation embryos, by using RNA fluorescence *in situ* hybridization (RNA-FISH), which allows nascent transcripts to be visualized. We found that *Oct4* is actively expressed in a small proportion of 2-cell-stage embryos, but most blastomeres express *Oct4* at the 4- and 8-cell stage (Fig. 1a, b). We detected active *Nanog* expression in about half of all 4- and 8-cell-stage blastomeres, which is consistent with previous reports of heterogeneous *Nanog* expression⁷. Surprisingly, although all of the other genes that we analysed were expressed biallelically, *Nanog* invariably showed monoallelic expression in 2-, 4- and 8-cell-stage embryos (Fig. 1a, b and Supplementary Movie 1). We next addressed whether the monoallelic

expression of *Nanog* is random or imprinted, by using allele-specific, single-cell PCR with reverse transcription (RT-PCR) on 8-cell-stage embryos from C57BL/6J \times *Mus musculus castaneus* crosses. Consistent with our RNA-FISH results, most blastomeres expressed *Nanog* monoallelically (Fig. 1c and Supplementary Fig. 1). We observed no bias in *Nanog* expression from either the C57BL/6J or the *M. musculus castaneus* allele (Fig. 1c). We conclude that *Nanog* expression is monoallelic and random during early pre-implantation development.

Nanog expression becomes restricted to the inner cell mass (ICM) of the blastocyst^{1,2}, where it is needed for the naive epiblast to attain pluripotency⁶. As expected, the proportion of cells expressing *Nanog* and *Oct4* declined gradually as the expression of these genes became restricted to inner cells (Fig. 2a). Remarkably, the proportion of biallelism in *Nanog*-expressing cells increased progressively from 2% at the 8-cell stage to 70% of the inner cells in the late blastocyst (Fig. 2b and Supplementary Movie 2). This switch to biallelism of *Nanog* is in contrast to the consistent biallelic expression of *Oct4* and *Actb* (Fig. 2a). Sequential RNA-FISH and DNA-FISH confirmed the biallelic expression of *Nanog* (Supplementary Fig. 1). The ICM of the late blastocyst contains two lineages: the extra-embryonic primitive endoderm, and the ‘ground-state’ pluripotent epiblast^{6,8}, which gives rise to the embryo. Inner cells expressing *Nanog* biallelically also express *Oct4* but not *Gata4*, a primitive endoderm marker⁹, and therefore are epiblast cells (Fig. 2c). This finding suggests that *Nanog* is predominantly expressed from both alleles in the naive epiblast. Furthermore, *Nanog* expression gradually reversed to monoallelic expression after implantation (Supplementary Fig. 2), in line with the downregulation of *Nanog* and concomitant loss of

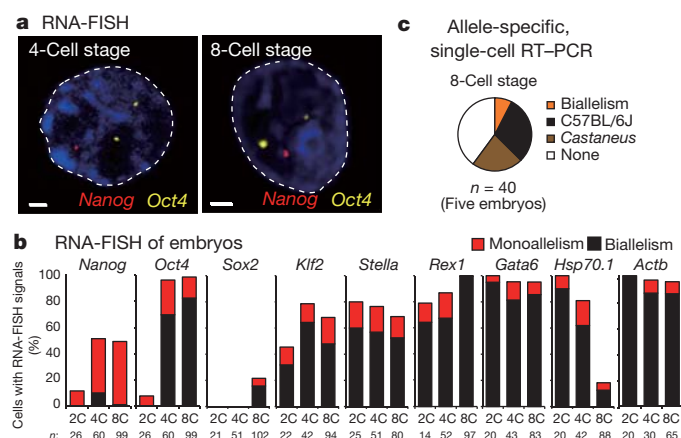


Figure 1 | *Nanog* expression is monoallelic in early embryos. **a**, RNA-FISH for *Nanog* and *Oct4* in 4- and 8-cell-stage nuclei. Scale bar, 2 μ m. **b**, Proportion of allelic expression for each gene in 2-cell-stage (2C), 4-cell-stage (4C) and 8-cell-stage (8C) blastomeres. *n*, number of cells analysed. **c**, Summary of single-cell RT-PCR data for *Nanog* in 8-cell-stage blastomeres, according to biallelic, monoallelic (C57BL/6J or *M. musculus castaneus* (*Castaneus*)) and no expression (none).

¹Institut de Génétique et de Biologie Moléculaire et Cellulaire, CNRS/INSERM U964, Université de Strasbourg, F-67404 Illkirch, Cité Universitaire de Strasbourg, France.

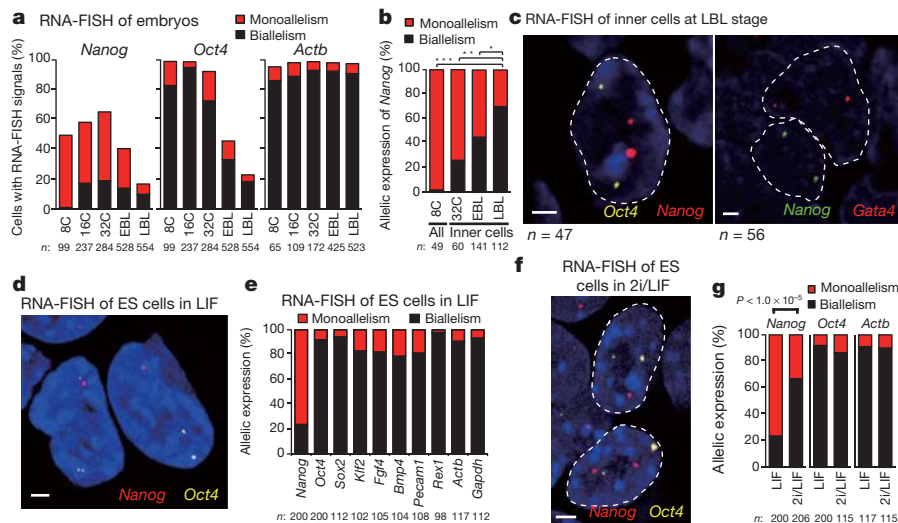


Figure 2 | *Nanog* shows biallelic expression in the naive, pluripotent epiblast. **a**, RNA-FISH for *Nanog*, *Oct4* and *Actb* in 8-cell-stage (8C), 16-cell-stage (16C) and 32-cell-stage (32C) embryos (before cavitation), and early blastocysts (EBLs; E3.5) and late blastocysts (LBLs; E4.25). Data are shown as the percentage of *Nanog*-expressing cells (or *Oct4*- or *Actb*-expressing cells) relative to all cells analysed. **b**, Quantification of cells showing monoallelic or biallelic *Nanog* expression in *Nanog*-expressing 8-cell-stage blastomeres and inner cells at the 32-cell, EBL and LBL stages. *, $P = 8.8 \times 10^{-5}$; **, $P = 3.1 \times 10^{-3}$; ***, $P < 1.0 \times 10^{-13}$ (Fisher's exact test). **c**, RNA-FISH in

ICM cells of the LBL. The dashed line delineates individual ICM cells. **d**, ES cells cultured with serum and LIF display primarily monoallelic *Nanog* expression. A representative RNA-FISH image of ES cells is shown. **e**, Distribution of monoallelic and biallelic expression of *Nanog* and the indicated genes in ES cells cultured with LIF. **f**, A representative RNA-FISH image of nuclei from ES cells cultured with 2i/LIF, showing biallelic expression of *Nanog*. **g**, Treatment of ES cells with 2i/LIF leads to increased biallelic *Nanog* expression. The proportion of allelic expression of *Nanog*, *Oct4* and *Actb* is shown for ES cells cultured with LIF or 2i/LIF. **a**, **b**, **g**, n , number of cells analysed. **c**, **d**, **f**, Scale bars, 2 μ m.

pluripotency². Thus, *Nanog* expression switches transiently from monoallelic to biallelic during the formation of the pluripotent epiblast *in vivo*, coinciding with the completion of the reprogramming process.

We next investigated the allelic expression of *Nanog* in ES cells. When cultured in medium containing serum and LIF, 60–70% of ES cells express *Nanog* (refs 10, 11) (Supplementary Fig. 4). Under these conditions, *Nanog* is predominantly expressed from a single allele (Fig. 2d, e and Supplementary Movie 3). By contrast, all of the other genes analysed showed biallelic expression, including other pluripotency-associated genes (*Oct4*, *Sox2*, *Klf2* and *Fgf4*), heterogeneously expressed genes (*Pecam1*, *Bmp4* and *Rex1* (also known as *Zfp42*)) and housekeeping genes (*Actb* and *Gapdh*) (Fig. 2d, e). Allele-specific, single-cell RT-PCR with hybrid ES cells showed that *Nanog* expression is independent of parental origin, similarly to the early embryo (Supplementary Fig. 3). Thus, *Nanog* shows random monoallelic expression both in early pre-implantation embryos and in ES cells.

The proportion of monoallelic versus biallelic *Nanog* expression in the naive epiblast is significantly different from that in ES cells, the latter showing much lower biallelic expression (Fig. 2b, e). We hypothesized that this could reflect the more homogeneous expression of *Nanog* in the naive epiblast compared to ES cells^{6,8}. Ground-state pluripotency can be established *in vitro* through pharmacological inhibition (with two inhibitors (2i)) of MEK and GSK3 β ¹². ES cells cultured with 2i show increased levels of *Nanog* messenger RNA^{6,12} and homogeneous distribution of NANOG protein¹³ (Supplementary Fig. 4). Strikingly, we found that culturing ES cells in 2i resulted in a significant increase in biallelic *Nanog* expression (Fig. 2f, g, Supplementary Fig. 3 and Supplementary Movie 4). We confirmed *Nanog* biallelic expression with sequential RNA-FISH and DNA-FISH (data not shown). To address which signalling pathway(s) regulates the allelic expression of *Nanog*, we used combinations of pharmacological inhibitors^{6,12,14}. We found that blocking fibroblast growth factor (FGF)–extracellular regulated kinase (ERK) signalling was sufficient to trigger the biallelic expression of *Nanog* (Supplementary Fig. 4). The naive epiblast and ES cells are thought to be maintained in a ground state by preventing the differentiation that is induced through FGF–ERK signalling^{8,12}. Thus, collectively these results suggest that ground-state pluripotency might be achieved through the activation of the second allele of *Nanog*.

The above changes in *Nanog* allelic expression upon pharmacological treatment suggest that the *Nanog* alleles are dynamically regulated. To visualize the dynamics of activation of the two *Nanog* alleles in individual cells, we generated an ES-cell reporter in which a gene encoding destabilized Turbo green fluorescent protein (TurboGFP) was inserted immediately downstream of the NANOG-coding region in one *Nanog* allele and a gene encoding destabilized mCherry was inserted similarly into the other *Nanog* allele (Fig. 3a and Supplementary Fig. 5). Both fluorescent proteins dissociate from NANOG by self-cleavage of a 2A peptide and do not alter NANOG function. We observed heterogeneous distribution of TurboGFP or mCherry in ES cells grown in LIF, confirming that the monoallelic expression of *Nanog* is random (Fig. 3b). This is in contrast to the homogeneous distribution in ES cells grown in 2i, in which most cells expressed both fluorescent proteins, reflecting biallelic expression of *Nanog* (Fig. 3b). Using time-lapse analysis, we observed dynamic fluctuation of *Nanog* expression, in agreement with previous reports^{10,11} (Supplementary Fig. 6 and Supplementary Movie 5). Moreover, we found that *Nanog* expression can switch between alleles. In most cases, the allelic switch occurred through an intermediate state of biallelic *Nanog* expression or no *Nanog* expression over several cell cycles. In a minority of cases (2%), rapid allelic switching occurred during a single cell cycle (Fig. 3c and Supplementary Fig. 6). This switching is in contrast to other monoallelic genes, including imprinted genes, for which the inactive status of an allele is stably maintained¹⁵.

Given the potential importance of activation of the second *Nanog* allele in establishing ground-state pluripotency, we next addressed the chromatin signatures underlying biallelic activation. We asked whether the allelic expression of *Nanog* obeys any epigenetic feature. We found that *Nanog* replicated asymmetrically in ES cells cultured with LIF alone—a feature of monoallelically expressed genes—but changed its replication pattern towards symmetric replication in ES cells cultured with 2i and LIF (denoted 2i/LIF) (Fig. 3d, e). This pattern is in contrast to the invariable asymmetric replication of imprinted genes (*Snurf* and *H19*; Fig. 3d, e) and is consistent with changes in allelic *Nanog* expression. Replication timing is a distinctive epigenetic fingerprint^{5,16,17}, thus suggesting that the allelic regulation of *Nanog* could be associated with epigenetic signatures. Indeed, monoallelic expression is accompanied

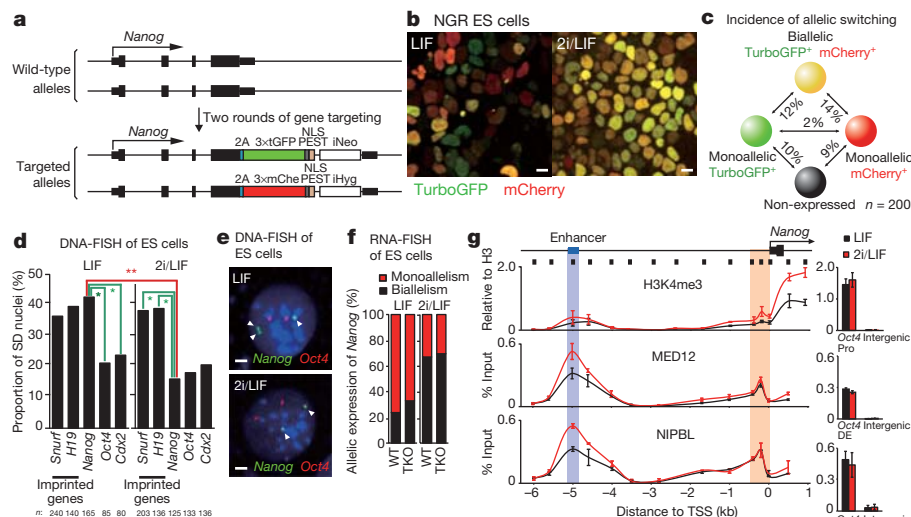


Figure 3 | The switch to biallelic expression of *Nanog* is accompanied by changes in replication timing and increased recruitment of mediator and cohesin. **a**, Schematic of the *Nanog* knock-in reporter NGR. A PEST motif in the carboxy terminus of the fluorescent proteins allows monitoring of dynamic *Nanog* expression. iHyg, internal ribosome entry site (IRES) hygromycin; iNeo, IRES neomycin; mChe, mCherry; NLS, nuclear localization signal; tGFP, TurboGFP. **b**, Representative image of NGR ES cells cultured with LIF or 2i/LIF. Scale bar, 10 μ m. **c**, The incidence of allelic switching of *Nanog* expression in ES cells. Cells were classified into four groups: monoallelic (TurboGFP-positive, green), monoallelic (mCherry-positive, red), biallelic (TurboGFP- and mCherry-positive, yellow) and no expression (black). The proportion of cells undergoing a transition between these four groups during a single cell cycle is indicated. Overall, 47% of cells showed a colour change in this period. n , number of cells analysed. **d**, The asymmetric replication of *Nanog* in ES cells cultured with LIF changes to symmetric replication upon treatment with 2i.

by allele-specific DNA methylation and/or histone modifications^{15,18}. However, we found that ES cells lacking all three DNA methyltransferases¹⁹ had similar *Nanog* allelic expression to wild-type ES cells (Fig. 3f), indicating that the allelic regulation of *Nanog* is independent of DNA methylation, unlike that of imprinted genes¹⁵.

Next, we assessed allelic histone modifications on *Nanog* by using chromatin immunoprecipitation (ChIP), taking advantage of the fact that ES cells grown in LIF express *Nanog* mainly in a monoallelic manner and those grown in 2i/LIF primarily have two active alleles. ChIP showed a significantly higher enrichment of H3K4me3, a modification of active chromatin, at the transcription start site (TSS) of *Nanog* in ES cells grown in 2i/LIF than those grown in LIF alone (Fig. 3g). This increase is in line with the higher biallelic expression in 2i/LIF than LIF alone. The H3K4me3 enrichment at the *Nanog* TSS was approximately twofold in 2i/LIF, possibly suggesting that H3K4me3 is enriched at the active *Nanog* allele but absent at the inactive one in ES cells expressing *Nanog* monoallelically. We observed significantly higher recruitment of the mediator subunit MED12, as well as the cohesin-loading factor NIPBL, to the *Nanog* enhancer in ES cells grown in 2i/LIF than in ES cells grown in LIF alone (Fig. 3g). This increased recruitment was accompanied by increased RNA polymerase II initiation and elongation upon 2i treatment (Supplementary Fig. 7). Indeed, changes in chromatin architecture mediated through simultaneous binding of mediator and cohesin to the *Nanog* locus facilitate looping between the enhancer and promoter, resulting in *Nanog* gene activation²⁰. Collectively, our data suggest that *Nanog* allelic expression might be actively regulated through chromosomal changes over the *Nanog* locus.

Mice that are heterozygous for *Nanog* (*Nanog*^{+/-}) are born at normal Mendelian ratios¹. We hypothesized that if *Nanog* biallelic expression is important for the formation of the ICM and/or for the acquisition of ground-state pluripotency, then *Nanog*^{+/-} embryos may show developmental defects at the onset of *Nanog* biallelic

The cell nuclei were classified as single/double (SD), single/single (SS) and double/double (DD) according to DNA-FISH signals⁵. n , number of nuclei analysed. *, $P < 4 \times 10^{-7}$; **, $P < 1.4 \times 10^{-3}$ (Fisher's exact test).

e, Representative image of DNA-FISH for *Nanog* (arrowheads) and *Oct4* in ES cells cultured with LIF or 2i/LIF. Scale bar, 2 μ m. **f**, *Nanog* allelic expression is unaffected in the absence of DNA methyltransferase activity. Quantification of RNA-FISH for *Nanog* in wild-type (WT) ES cells and ES cells lacking all three DNA methyltransferases (TKO) cultured with LIF or 2i/LIF. **g**, ChIP for H3K4me3, MED12 or NIPBL along the *Nanog* locus (black line, top) in ES cells cultured with LIF or 2i/LIF. The position of the ChIP amplicons is depicted by the thick boxes below the line, the TSS by an arrow, the first exon by the black box on the line, and the distal enhancer by the blue box on the line. The *Oct4* promoter region (*Oct4* Pro) and distal enhancer (*Oct4* DE) were positive controls (right)²⁰. The mean \pm s.d. of three independent biological replicates is shown.

expression. To address this, we counted ICM cells in individual *Nanog*^{+/-} blastocysts by using an anti-OCT4 antibody. Freshly collected embryonic day 3.5 (E3.5) *Nanog*^{+/-} and *Nanog*^{+/+} blastocysts had similar ICM cell numbers (Supplementary Fig. 8a). This finding is in agreement with previous reports^{6,21} and with our observations documenting *Nanog* monoallelic expression before this embryonic stage. By contrast, we found significantly fewer ICM cells in *Nanog*^{+/-} blastocysts as the embryo progressed through development to form the nascent epiblast (Fig. 4a), which also showed increased apoptosis in the ICM (Supplementary Fig. 8b). No significant difference in the number of NANOG-positive cells was observed between wild-type (*Nanog*^{+/+}) and heterozygous (*Nanog*^{+/-}) embryos (Fig. 4b, c). However, *Nanog*^{+/-} embryos showed delayed primitive endoderm formation (Fig. 4b, c). Indeed, primitive endoderm formation depends on a functional epiblast in a non-cell-autonomous manner²¹, suggesting that the epiblast in *Nanog*^{+/-} embryos is functionally impaired. It should be noted that the reduction in primitive endoderm cells does not fatally compromise development, owing to the regulatory capacity of the early post-implantation embryo. Thus, we conclude that the biallelic expression of *Nanog* is necessary for the timely maturation of the ICM into a functional epiblast and the accompanying primitive endoderm.

NANOG has a dose-sensitive action in reprogramming⁶. Overexpression of NANOG is sufficient to prevent ES-cell differentiation²; conversely, *Nanog*^{+/-} ES cells show spontaneous differentiation^{22,23}, suggesting that the tight regulation of *Nanog* levels is crucial for both reprogramming and differentiation. We found that this dose regulation occurs principally at the allelic level. Monoallelic expression is predominant during the early reprogramming phase in cleavage stages, but it switches to biallelic expression as the ICM matures into the pluripotent epiblast (Fig. 4d). Monoallelic *Nanog* expression in ES cells may be representative of the late ICM or the early post-implantation

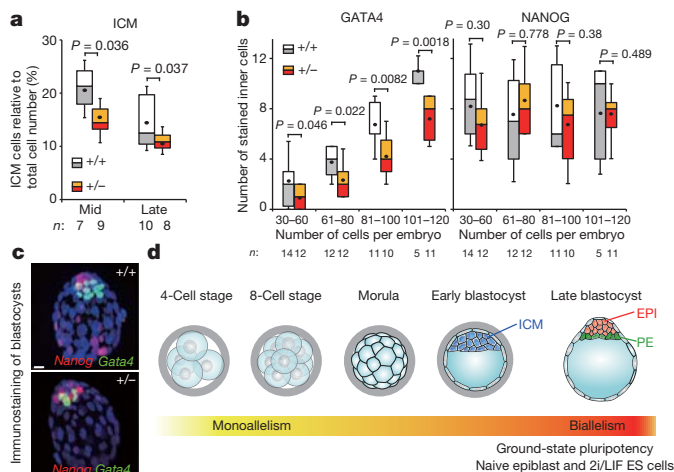


Figure 4 | Biallelic expression of *Nanog* is required for the timely maturation of the ICM derivatives. **a**, Reduced proportion of ICM derivatives in *Nanog*^{+/-} blastocysts compared to *Nanog*^{+/+} blastocysts. Mid and late blastocysts were staged according to their total cell number. Data are expressed as the percentage of ICM cells per embryo. The box plot represents the ninetieth, seventy-fifth, fiftieth (median), twenty-fifth and tenth percentiles, as well as the mean (black dot). **b**, Delayed primitive endoderm (PE) formation in *Nanog*^{+/-} blastocysts. Embryos were staged in groups according to their total cell number. The number of PE (GATA4-positive) and epiblast (EPI; NANOG-positive) cells was determined. **c**, Representative maximal confocal projections of late blastocysts, showing a reduced proportion of ICM derivatives in *Nanog*^{+/-} embryos. Scale bar, 10 μ m. **d**, Model for the allelic regulation of *Nanog* and the acquisition of ground-state pluripotency. *Nanog* is expressed monoallelically during the cleavage stages and up to the early blastocyst. Biallelic expression of *Nanog* accompanies the formation of the fully reprogrammed, pluripotent, epiblast. Biallelic expression of *Nanog* also occurs *in vitro* under conditions that mimic the ground-state pluripotency of the naive epiblast (ES cells cultured with 2i/LIF).

epiblast, in which there is a switch back to monoallelic expression. The expression of *Nanog* can switch between alleles, unlike most known monoallelic genes, raising the possibility that the regulation of *Nanog* allelic expression could be unique. Our evidence that such regulation is independent of DNA methylation and does not seem to involve typical repressive modifications (H3K27me₃, H3K9me₁, H3K9me₂ and H3K9me₃; data not shown) also supports this possibility. The ground-state pluripotency is facilitated by the suppression of FGF-ERK signalling^{8,12}. Our work suggests that the allelic regulation of *Nanog* is involved in this process and that the increased expression of *Nanog* that accompanies the establishment of ground-state pluripotency is achieved through activation of its second allele. The dose regulation of *Nanog* through allelic switching could also contribute to promoting heterogeneity, which could provide a window of opportunity both for establishing ground-state pluripotency in the epiblast and for lineage segregation towards the primitive endoderm. We propose that the regulation of *Nanog* at the allelic level provides a novel mechanism to establish ground-state pluripotency in the naive epiblast. Thus, our work has uncovered an unexpected role for the allelic regulation of key transcription factors, as a novel mechanism in stabilizing pluripotency.

METHODS SUMMARY

The embryos were from F₁ (C57BL/6J \times CBA/H) or C57BL/6J \times *M. musculus castaneus* crosses. The blastocysts from *Nanog*^{Gt1/+} \times C57BL/6J crosses²¹ were genotyped after immunostaining (47 *Nanog*^{+/+} and 49 *Nanog*^{+/-} were analysed). The ES-cell lines and culture conditions are described in the Methods. Briefly, when referring to LIF, ES cells were cultured in 15% FCS and LIF; when referring to 2i/LIF, ES cells were cultured in 15% FCS, LIF, 3 μ M CHIR99021 and 1 μ M PD0325901. Embryos and ES cells were fixed with 4% paraformaldehyde in PBS for immunostaining. Images were acquired on a TCS SP5 confocal microscope (Leica). Image analysis was performed with the programs ImageJ, LAS AF Lite (Leica) or Imaris (Bitplane). For single-cell RT-PCR analysis, individual 8-cell-stage

blastomeres were mechanically dissociated, and the cells were collected in 3.1 μ l Cell Lysis Buffer (Ambion) and processed for RT-PCR then digested with *Av*II. All primer sequences are listed in Supplementary Table 1.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 22 June; accepted 19 December 2011.

Published online 12 February 2012.

- Mitsui, K. *et al.* The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell* **113**, 631–642 (2003).
- Chambers, I. *et al.* Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell* **113**, 643–655 (2003).
- Nichols, J. *et al.* Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* **95**, 379–391 (1998).
- Reik, W. Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature* **447**, 425–432 (2007).
- Gribnau, J., Hochedlinger, K., Hata, K., Li, E. & Jaenisch, R. Asynchronous replication timing of imprinted loci is independent of DNA methylation, but consistent with differential subnuclear localization. *Genes Dev.* **17**, 759–773 (2003).
- Silva, J. *et al.* Nanog is the gateway to the pluripotent ground state. *Cell* **138**, 722–737 (2009).
- Dietrich, J. E. & Hiiragi, T. Stochastic patterning in the mouse pre-implantation embryo. *Development* **134**, 4219–4231 (2007).
- Nichols, J., Silva, J., Roode, M. & Smith, A. Suppression of Erk signalling promotes ground state pluripotency in the mouse embryo. *Development* **136**, 3215–3222 (2009).
- Morrissey, E. E., Ip, H. S., Lu, M. M. & Parmacek, M. S. GATA-6: a zinc finger transcription factor that is expressed in multiple cell lineages derived from lateral mesoderm. *Dev. Biol.* **177**, 309–322 (1996).
- Chambers, I. *et al.* Nanog safeguards pluripotency and mediates germline development. *Nature* **450**, 1230–1234 (2007).
- Kalmar, T. *et al.* Regulated fluctuations in Nanog expression mediate cell fate decisions in embryonic stem cells. *PLoS Biol.* **7**, e1000149 (2009).
- Ying, Q. L. *et al.* The ground state of embryonic stem cell self-renewal. *Nature* **453**, 519–523 (2008).
- Wray, J., Kalkan, T. & Smith, A. G. The ground state of pluripotency. *Biochem. Soc. Trans.* **38**, 1027–1032 (2010).
- Canham, M. A., Sharov, A. A., Ko, M. S. & Brickman, J. M. Functional heterogeneity of embryonic stem cells revealed through translational amplification of an early endodermal transcript. *PLoS Biol.* **8**, e1000379 (2010).
- Weaver, J. R., Susiarjo, M. & Bartolomei, M. S. Imprinting and epigenetic changes in the early embryo. *Mamm. Genome* **20**, 532–543 (2009).
- Shufaro, Y. *et al.* Reprogramming of DNA replication timing. *Stem Cells* **28**, 443–449 (2010).
- Jørgensen, H. F. *et al.* The impact of chromatin modifiers on the timing of locus replication in mouse embryonic stem cells. *Genome Biol.* **8**, R169 (2007).
- Magklara, A. *et al.* An epigenetic signature for monoallelic olfactory receptor expression. *Cell* **145**, 555–570 (2011).
- Tsumura, A. *et al.* Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. *Genes Cells* **11**, 805–814 (2006).
- Kagey, M. H. *et al.* Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **467**, 430–435 (2010).
- Messerschmidt, D. M. & Kemler, R. Nanog is required for primitive endoderm formation through a non-cell autonomous mechanism. *Dev. Biol.* **344**, 129–137 (2010).
- Hough, S. R., Clements, I., Welch, P. J. & Wiederholt, K. A. Differentiation of mouse embryonic stem cells after RNA interference-mediated silencing of OCT4 and Nanog. *Stem Cells* **24**, 1467–1475 (2006).
- Hatano, S. Y. *et al.* Pluripotential competence of cells associated with Nanog activity. *Mech. Dev.* **122**, 67–79 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. Okano for providing the triple DNA methyltransferase knockout cells, P. Avner and P. Clerc for hybrid ES cells, H. Schöler for the *Oct4* probe (GOF6.1), J. Jaubert for *M. musculus castaneus* mice, R. Kemler and S. Rudloff for the *Nanog*^{Gt1} mice, G. Charvin for advice on time-lapse analysis. We also thank M. Koch, A. Dierich, M.-C. Birling, E. Heard and I. Okamoto for advice, and A. Burton, E. Heard and O. Pourquie for critical reading of the manuscript. M.-E.T.-P. acknowledges funding from AVENIR/INSERM, ANR-09-Blanc-0114, Epigenesis NoE and FRM Alsace. Y.M. is a recipient of an EMBO long-term fellowship (ALTF864-2008, 2009) and a JSPS postdoctoral fellowship (2010–2011).

Author Contributions Y.M. conceived, designed and performed the experiments in this study and analysed the data; M.-E.T.-P. conceived the project and designed and supervised the study. Y.M. and M.-E. T.-P. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.-E.T.-P. (metp@igbmc.fr).

METHODS

Embryo collection and culture. Embryos were collected from ~6 week-old F₁ (C57BL/6J × CBA/H) superovulated females crossed with F₁ males. To obtain F₁ hybrid embryos, superovulated C57BL/6J females were crossed with *M. musculus castaneus* males. Embryos were collected at the following times after human chorionic gonadotrophin injection: the 2-cell stage (48 h), 4-cell stage (54 h), 8-cell stage (68 h), 16-cell stage (80 h), 32-cell stage (90 h), early blastocyst stage (98 h) and late blastocyst stage (114 h). To obtain late blastocysts, freshly collected early blastocysts were cultured in KSOM for 16 h. To obtain E4.5, E4.75 and E5.0 stage embryos, embryos were collected from naturally mated females. The analysis of *Nanog*^{+/-} embryos was done using the *Nanog*^{Gt1} line²¹. Blastocysts from *Nanog*^{+/-} × C57BL/6J (+/+) crosses were processed for immunostaining and reconstructed in three dimensions using Imaris software. Individual ICM cells were counted after immunostaining with an anti-OCT4 antibody. The number of primitive endoderm and epiblast cells was determined by counting individual cells after immunostaining for GATA4 and NANOG. The total number of cells was determined by counting nuclei stained with 4',6-diamidino-2-phenylindole (DAPI). Individual blastocysts were genotyped after confocal acquisition as described previously²¹. Primer sequences for genotyping are listed in Supplementary Table 1. All experiments were performed in accordance with the current legislation in France and the approval of the Regional Ethics Committee (ComEth's).

ES-cell culture. Mouse ES-cell lines, E14, BC1 and TKO (the latter of which lacked all three DNA methyltransferases), were cultured in DMEM with GlutaMAX (Invitrogen) containing 15% FCS, LIF, 1 mM sodium pyruvate, penicillin/streptomycin and 0.1 mM 2-mercaptoethanol. The treatment of ES cells with inhibitors was performed using 3 μM CHIR99021 (a GSK3β inhibitor), 1 μM PD0325901 (a MEK inhibitor), 5 μM PD173074 (an FGF receptor tyrosine kinase activity inhibitor) and 3 μM PD184352 (a MEK inhibitor). The hybrid ES-cell line (BC1) was derived from E3.5 embryos from C57BL/6 × *M. musculus castaneus* crosses as described previously¹².

Gene targeting. Mouse ES cells from the BD10 line (C57BL/6) were electroporated with a first targeting vector consisting of 2A-3×TurboGFP-NLS-PEST-IRES-Neo. The resultant G418-resistant ES-cell (*Nanog*^{GFP/+}) clone was subsequently subjected to a second targeting with a vector containing 2A-3×mCherry-NLS-PEST-IRES-Hyg. Genomic DNAs from both G418- and hygromycin-B-resistant colonies were screened for homologous recombination into each *Nanog* allele by PCR and Southern blotting.

Immunostaining. Embryos or ES cells were fixed with 4% paraformaldehyde in PBS for 20 min at room temperature. After washing with PBS, embryos were permeabilized with 0.5% Triton X-100 in PBS for 10 min and then incubated in blocking solution (0.2% BSA in PBS) for 30 min. Primary antibodies were anti-OCT4 (611202, BD Pharmingen), anti-NANOG (MLC-51, eBioscience) and anti-GATA4 (sc-1237, Santa Cruz Biotechnology). After incubation in blocking solution containing primary antibodies for 1 h, cells or embryos were washed three times with 0.01% Triton X-100 in PBS for 5 min each and then incubated in blocking solution containing secondary antibodies labelled with Alexa fluorophores (Invitrogen), Cy3, Cy5 (Jackson ImmunoResearch) or Kodak X-SIGHT 640 (Carestream Health). After washing with PBS, mounting was done in VECTASHIELD (Vector Labs). Images were collected on a TCS SP5 confocal microscope (Leica). For all images, DAPI is shown in blue. Image analysis was performed using the software ImageJ, LAS AF Lite (Leica) and Imaris (Bitplane).

Whole mount RNA-FISH. After removal of the zona pellucida by incubation in acidic Tyrode's solution (Sigma), embryos were washed twice with PBS and then incubated in fixative containing 4% paraformaldehyde/1×PBS for 20 min at room temperature. Embryos were permeabilized with 0.5% Triton X-100 in fixative for 10 min. After washing with PBS three times, embryos were pre-hybridized in a 2 μl drop of hybridization buffer covered with mineral oil on a 35-mm glass bottom dish for 30 min at 50 °C. The hybridization buffer consisted of 50% formamide, 10% dextran sulphate, 2×SSC, 1 μg μl⁻¹ Cot1 DNA, 1 μg μl⁻¹ yeast tRNA (Roche), 2 mM vanadyl ribonucleoside complex (New England BioLabs), 1 mg ml⁻¹ polyvinyl pyrrolidone (PVP), 1 mM EDTA, 0.1% Triton X-100 and 1 mg ml⁻¹ BSA. Embryos were then transferred to a 2 μl drop of hybridization buffer containing 5 ng μl⁻¹ fluorescent probes and incubated at 50 °C overnight. After washing three times with 2×SSC, 0.1% Triton X-100 and 1 mg ml⁻¹ PVP at 50 °C for 10 min, embryos were then subjected to a gradient series (20%, 40%, 60%, 80% and 100%) of VECTASHIELD containing DAPI. Images were acquired on a TCS SP5 confocal microscope with a 63× glycerol immersion objective lens. For RNA-FISH of ES cells, ES cells grown on gelatin-coated coverslips were fixed and permeabilized as described above. After briefly washing with PBS, the coverslip was dehydrated in 70% ethanol and 100% ethanol for 10 min each and was subsequently air dried. Hybridization buffer containing the corresponding probes was applied to the coverslip. Subsequent hybridization, washing and mounting were

performed as described above. The plasmids, BACs and fosmids used as probes for RNA-FISH were as follows: PBSII-Na1234 for *Nanog*, GOF6.1 for *Oct4*, G135P62831G1 and G135P630G2 for *Gata4*, RP23-476J7 or G135P6579H5 for *Actb*, G135P60852B4 for *Sox2*, G135P65370D8 for *Klf2*, G135P600983F2 for *Fgf4*, G135P60839B3 for *Bmp4*, RP23-129L1 for *Gata6*, G135P65026G5 for *Hsp70.1*, G135P604053D8 for *Pecam1*, G135P69364G7 for *Stella*, G135P602899G3 for *Rex1* and G135P60163E2 for *Gapdh*. The PBSII-Na1234 plasmid contains the NANOG-coding region without any repetitive sequences, as determined by RepeatMasker. BAC and fosmid probes were obtained from the BACPAC Resources Center. The specificity of all probes was confirmed by PCR using specific primers. Probes were labelled with handmade Alexa 488-dATP, TAMRA-dATP, ATTO 647N-dATP using a nick translation kit (Roche) and purified with a QIAquick PCR Purification Kit (QIAGEN). Images were analysed by the software ImageJ and LAS AF Lite. For all of the images, DAPI is shown in blue. It should be noted that the monoallelic expression of *Nanog* was independent of the genetic background, as it was reproducible in all of the mouse strains analysed (F₁ (C57BL/6J × CBA/H), F₁ (C57BL/6J × *M. musculus castaneus*), CD1 and C57BL/6J; data not shown).

Statistical analysis of RNA-FISH. Only interphase cells were taken into account for all of the analyses. Images were analysed across three-dimensional planes by using the software ImageJ and LAS AF Lite. The RNA-FISH signals, which were detected over three sequential Z-sections were judged as positive signals to distinguish background noise. Single RNA-FISH spots or two adjacent spots within 1 μm, which might result from DNA replication, were judged as monoallelism. Cells with two, three or four separate RNA-FISH spots were scored as biallelism. Statistical analysis was done using Fisher's exact test, and the number of cells analysed in all experiments is indicated in the corresponding figure.

DNA-FISH. DNA-FISH was performed as previously described²⁴. Probes for DNA-FISH were as follows: RP23-117I23 for *Nanog*, RP23-213M12 for *Oct4*, RP23-476J7 for *Actb*, RP23-106M4 for *Cdx2*, RP23-239M21 for *Snurf* and G135P602165C11 for *H19*. For all DNA-FISH images, DAPI is shown in blue.

Allele-specific, single-cell RT-PCR. Individual 8-cell-stage blastomeres from hybrid embryos or single hybrid BC1 ES cells were used for this assay. Embryos or ES cells were treated with 0.25% trypsin and 1 mM EDTA at 37 °C for 5 min with gentle pipetting. Single cells were manually collected by mouth pipette aided by a finely pulled glass tip, directly into 3.1 μl Cell Lysis Buffer (Ambion) containing RNase inhibitor (Invitrogen) and 73 nM specific primers for *Nanog* and *Actb*. The single-cell samples were snap frozen in liquid nitrogen and stored at -80 °C until use. Lysates were incubated at 65 °C for 5 min and placed on ice for 5 min. Reaction mixture (7 μl) (2 μl 5×RT buffer, 2 μl 2.5 mM dNTPs, 0.1 μl RNase inhibitor, 0.2 μl Transcriptor (Roche) and 2.7 μl water) was added to each tube. Reverse transcription was performed at 25 °C for 10 min, 37 °C for 15 min, 55 °C for 30 min and 85 °C for 5 min. The reaction mixtures were subjected to a first round of multiplex PCR with primer mixtures for *Nanog* and *Actb*. This first PCR reaction mixture (0.5 μl) was subsequently used for a second round of PCR with a specific primer set for *Nanog* or *Actb*. Pwo SuperYield DNA Polymerase (Roche) was used for PCR reactions. After the second PCR round, PCR products for *Nanog* were digested with the *AvaII* restriction enzyme and analysed by acrylamide gel electrophoresis. Primer sets for *Nanog*- and *Cdkn1*-flanking single nucleotide polymorphism (SNP) sites are listed in Supplementary Table 1. SNPs in the NANOG-coding region were identified by using The Jackson Laboratory's Mouse SNP database (<http://phenome.jax.org/SNP/>). SNPs in *Cdkn1* are described elsewhere²⁵. Primer sequences are listed in Supplementary Table 1.

Quantitative (real-time) PCR. RNA was extracted from ES cells with an RNeasy Mini Kit (QIAGEN) and reverse transcribed with Transcriptor (Roche). Real-time PCR was performed with SYBR Green JumpStart Taq ReadyMix (Sigma) on a LightCycler 480 Real-Time PCR System (Roche). The relative expression level of each gene was normalized to the *Gapdh* expression level. Primer sequences are listed in Supplementary Table 1.

ChIP assay. ChIP assays were performed with ES cells as previously described²⁶. Primer sequences are listed in Supplementary Table 1. The antibodies used were anti-histone H3 (ab1791, Abcam), anti-H3K4me3 (pAV-MEHAHS-024, Diagenode) anti-MED12 (A300-774A, Bethyl Laboratories), anti-NIPBL (301-779A, Bethyl Laboratories), anti-RNA polymerase II (PolII) (N-20, Santa Cruz Biotechnology), anti-PolII CTD Ser2P (ab5095, Abcam) and anti-PolII CTD Ser5P (ab5131, Abcam). The enrichment of histone modifications was quantified by real-time PCR as described above and normalized to histone H3. For MED12, NIPBL, PolII, PolII CTD Ser2P and PolII CTD Ser5P, the percentage input was calculated for each ChIP fraction.

Time-lapse imaging. NGR ES cells were cultured in Knockout DMEM (Invitrogen) supplemented with 15% Knockout Serum Replacement (Invitrogen), LIF, non-essential amino acids, glutamine, gentamicin and 2-mercaptoethanol, under conditions of 5% CO₂ and 95% air. The cells were grown on Culture-Insert (Ibidi) placed

on laminin-511 (Biolamina)-coated 3-cm glass-bottomed dishes (Mattek). The dishes were placed in an incubation chamber (Tokai Hit) on the microscope stage at 37 °C. An inverted microscope (Leica) attached to a Nipkow disk confocal microscope (Yokogawa Electric) and EMCCD camera (iXon, Andor Technology) was controlled with iQ software (Andor Technology). TurboGFP and mCherry were excited with 488-nm and 560-nm lasers, respectively. Images for each colour (green or red) were acquired across 17.5 µm (7 Z-planes) every 20 min for ~50 h.

Analysis of live-cell imaging data. Time-lapse data were analysed using the ImageJ software plug-in Circadian Gene Expression (CGE)²⁷. Fluorescent images were merged and used for the segmentation of nuclei. Cells ($n = 200$) were randomly selected, and the fluorescence intensities for TurboGFP and mCherry in each nucleus were measured in each time frame. Background intensity was subtracted from the mean fluorescence intensity and plotted as the expression level against time during a single cell cycle (~10 h) as shown in Supplementary Fig. 6. Cells that

were positive for TurboGFP, positive for mCherry, positive for both or negative for both were scored as monoallelic (green), monoallelic (red), biallelic (yellow) or non-expressed (black), respectively. Based on the kinetics of TurboGFP and mCherry during a single cell-cycle, the proportion of cells showing a transition between these four groups was calculated.

24. Bolzer, A. *et al.* Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. *PLoS Biol.* **3**, e157 (2005).
25. Lewis, A. *et al.* Epigenetic dynamics of the *Kcnq1* imprinted domain in the early embryo. *Development* **133**, 4203–4210 (2006).
26. Boyer, L. A. *et al.* Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122**, 947–956 (2005).
27. Sage, D., Unser, M., Salmon, P. & Dibner, C. A software solution for recording circadian oscillator features in time-lapse live cell microscopy. *Cell Div.* **5**, 17 (2010).

The microRNA miR-34 modulates ageing and neurodegeneration in *Drosophila*

Nan Liu^{1†}, Michael Landreh^{1†}, Kajia Cao^{2,3}, Masashi Abe¹, Gert-Jan Hendriks¹, Jason R. Kennerdell¹, Yongqing Zhu¹, Li-San Wang^{2,4,5} & Nancy M. Bonini^{1,6}

Human neurodegenerative diseases have the temporal hallmark of afflicting the elderly population. Ageing is one of the most prominent factors to influence disease onset and progression¹, yet little is known about the molecular pathways that connect these processes. To understand this connection it is necessary to identify the pathways that functionally integrate ageing, chronic maintenance of the brain and modulation of neurodegenerative disease. MicroRNAs (miRNA) are emerging as critical factors in gene regulation during development; however, their role in adult-onset, age-associated processes is only beginning to be revealed. Here we report that the conserved miRNA miR-34 regulates age-associated events and long-term brain integrity in *Drosophila*, providing a molecular link between ageing and neurodegeneration. Fly *mir-34* expression exhibits adult-onset, brain-enriched and age-modulated characteristics. Whereas *mir-34* loss triggers a gene profile of accelerated brain ageing, late-onset brain degeneration and a catastrophic decline in survival, *mir-34* upregulation extends median lifespan and mitigates neurodegeneration induced by human pathogenic polyglutamine disease protein. Some of the age-associated effects of miR-34 require adult-onset translational repression of *Eip74EF*, an essential ETS domain transcription factor involved in steroid hormone pathways. Our studies indicate that miRNA-dependent pathways may have an impact on adult-onset, age-associated events by silencing developmental genes that later have a deleterious influence on adult life cycle and disease, and highlight fly miR-34 as a key miRNA with a role in this process.

Recent evidence reveals that miRNA pathways are important in the adult nervous system, notably in the maintenance of neurons and in the regulation of genes and pathways associated with neurodegenerative disease^{2,3}. Given these findings, we considered that there may be a fundamental role for select miRNAs in ageing. We examined flies carrying a hypomorphic mutation in *loquacious* (*loqs*), a key gene in fly miRNA processing⁴ (Supplementary Fig. 1a). Flies bearing the *loqs*^{f00791} mutation were viable, but detailed examination indicated a significantly shortened lifespan (Supplementary Fig. 1b). Further analysis indicated that *loqs*^{f00791} flies showed late-onset brain morphological deterioration: although normal as young adults, by 25 days *loqs*^{f00791} flies developed large vacuoles in the retina and lamina of the brain (Supplementary Fig. 1c). Although developmental processes may contribute to shortened lifespan, the adult-onset brain degeneration of *loqs*^{f00791} mutants indicated that one or more specific miRNAs may be critically involved in age-associated events impacting on long-term brain integrity.

To explore this question, we determined whether specific miRNAs displayed age-modulated expression in the brain. RNA was isolated from dissected brains of adult flies of young (3 days), mid (30 days) and old time points (60 days). Using an array for *Drosophila* miRNAs, 29 were expressed in the adult brain (Fig. 1a). Whereas most miRNAs maintained a steady level or decreased with age, one miRNA, *mir-34*,

increased (Fig. 1a). Small RNA northern blot analysis confirmed that *mir-34* expression was barely detectable during development, but became high in the adult and was further upregulated with age (Supplementary Fig. 2a, b). Expression of *mir-34* was affected in *loqs*^{f00791} flies (Supplementary Fig. 1d). miR-34 falls into a category of *Drosophila* miRNAs whose processing requires the exoribonuclease *nibbler* (*nbr*)^{5,6}. In the adult, mature miR-34 displayed three major differentially sized forms (24 nucleotides, 22 nucleotides and 21 nucleotides) with a uniform 5' end, descending by single nucleotides at the 3' end which result from *nbr*-mediated trimming; only isoform c became upregulated with age (Supplementary Fig. 2c and Fig. 1b, c; see also refs 5–7).

miR-34 is a markedly conserved miRNA, with orthologues in fly, *Caenorhabditis elegans*, mouse and human showing identical seed sequence (Supplementary Fig. 2d). To define miR-34 function, flies deleted for the gene were generated (Supplementary Fig. 3a). The resulting *mir-34* mutant flies retained normal wild-type expression of neighbouring genes, but selectively lacked *mir-34* (Supplementary

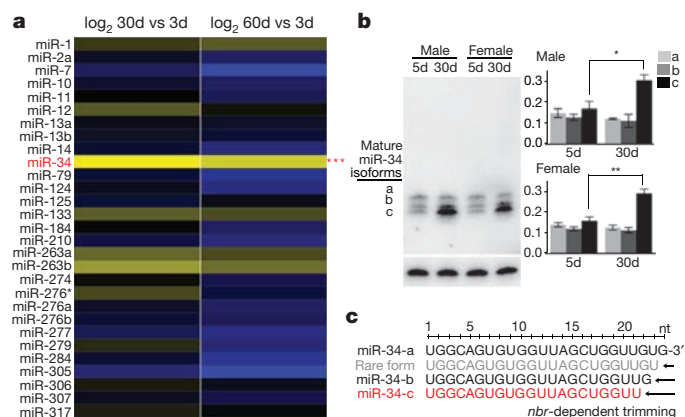


Figure 1 | *Drosophila* miR-34 expression is upregulated with age. **a**, Heat map of fold-change of *Drosophila* miRNAs in brains aged 3 days, 30 days and 60 days. Twenty-nine miRNAs (shown) were flagged present out of a total of seventy-eight. One-way analysis of variance defined significance for each miRNA over all time points ($***P < 0.001$; $n = 3$ replicates). Genotype: iso31. **b**, Fly miR-34 isoform c shows age-modulated expression in fly heads. Left panels: miR-34 shows three major mature forms (labelled a, b and c), but only isoform c increases with age. $n = 3$ independent experiments; signal density of all isoforms normalized at the same time point to 2S rRNA loading control. $*P < 0.01$; $**P < 0.001$; one-way analysis of variance, with post test: Tukey's multiple comparison test. Genotype: 5905. **c**, Sequences of miR-34 isoforms are generated through *nbr*-dependent 3'-end trimming.

¹Department of Biology, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. ²Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. ³Cell and Developmental Biology, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. ⁴Institute on Aging, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. ⁵Penn Center for Bioinformatics, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. ⁶Howard Hughes Medical Institute, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. [†]Present addresses: Division of Biological Sciences, Section of Neurobiology, Howard Hughes Medical Institute, University of California, San Diego, California 92093, USA (N.L.); Division of Physiological Chemistry I, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, SE-17177 Stockholm, Sweden (M.L.).

Fig. 3b, c). To interrogate age-associated phenotypes carefully, we generated *mir-34* null flies in the same uniform homogeneous genetic background (see Methods). *mir-34* mutants displayed no obvious developmental defects, consistent with its adult-onset expression. However, detailed examination of adult animals indicated that *mir-34* mutants, although showing normal adult appearance and early survival, displayed a catastrophic decline in viability just after 30 days (Fig. 2a and Supplementary Table 4). Analysis of age-associated functions revealed that young mutants (3 days) had normal locomotion and stress resistance, but by 20 days the mutants had dramatic climbing deficits and were markedly stress-sensitive compared to age-matched controls (Fig. 2b). Because *mir-34* expression was brain-enriched, we also examined the brain. Typically, older flies show sporadic, age-correlated vacuoles in the brain—a morphological hallmark of neural

deterioration⁸. *mir-34* mutants were born with normal brain morphology, but showed dramatic vacuolization with age, indicative of loss of brain integrity (Fig. 2c). Rescue with a 9-kb genomic DNA fragment containing *mir-34* and its endogenous *cis*-regulatory elements (Supplementary Fig. 3a, b) partially restored the age-associated expression of *mir-34* to *mir-34* null flies in the same homogeneous genetic background (Supplementary Fig. 3d). Although rescue was not complete, indicative of a complexity in genomic elements that regulate *mir-34*, rescue was sufficient to mitigate the mutant effects, indicating that miR-34 function normally underlies these age-associated aspects (Supplementary Table 1).

These data indicated that *mir-34* mutants were normal as young adults, but with age developed deficits reflective of much older animals, including loss of locomotion, stress sensitivity and brain deterioration,

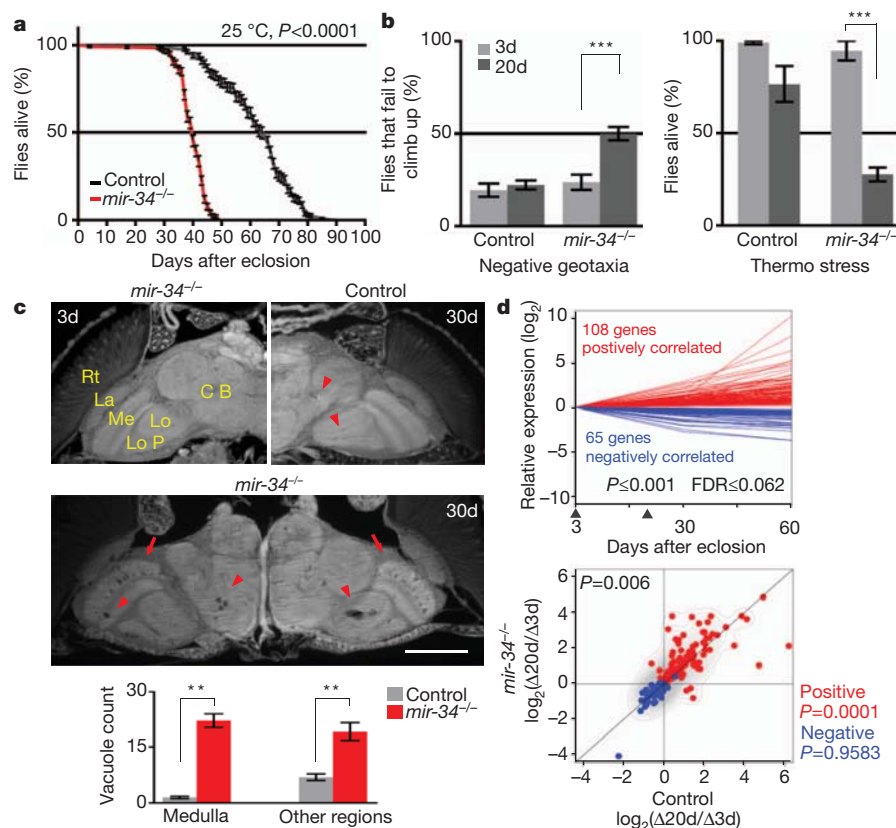


Figure 2 | miR-34 modulates age-associated processes. **a**, *mir-34* mutant flies have a shortened lifespan (control: 64 days median, 90 days maximal lifespan; *mir-34*: 40 days median, 64 days maximal lifespan; $P < 0.0001$, log-rank test). Mean \pm s.e.m., $n \geq 240$ male flies per curve. Genotypes: control, 5905; *mir-34*^{-/-}, *mir-34* null-1 in 5905 homogenous genetic background. **b**, *mir-34* mutant flies have late-onset behavioural deficits. Left: for locomotion behaviour, *mir-34* mutant flies show normal climbing at 3 days. At 20 days, $50 \pm 3.4\%$ *mir-34* mutant flies fail to climb; in contrast, only $22.1 \pm 2.4\%$ of control flies have defective climbing. Mean \pm s.e.m. of 3 experiments, $n = 120$ –140 male flies per experiment. Right panel: for stress resistance, *mir-34* mutant flies have normal resistance to heat stress at 3 days. *mir-34* mutant flies become markedly sensitive to heat shock with age, such that at 20 days, only $27.5 \pm 3.8\%$ survive after heat stress. In contrast, $76.7 \pm 9.6\%$ of control flies survive after the same treatment. Mean \pm s.e.m. of 3 experiments, $n = 120$ –140 male flies. $***P < 0.0001$ (two-way analysis of variance). Genotypes as in **a**. **c**, *mir-34* mutant flies show age-associated brain degeneration. Top-left panel: *mir-34* mutant flies have normal brain morphology at 3 days. Major anatomical structures: CB, central brain; La, lamina; Lo, lobula; LoP, lobula plate; Me, medulla; Rt, retina. At 3 days, control flies have normal brain morphology (not shown), but develop a small number of sporadic vacuoles at 30 days (top-right panel, arrowheads). Middle panel: aged *mir-34* mutants (30 days) show striking vacuoles in the medulla (arrows) and other regions of

the brain (arrowheads). Bottom: the number of vacuoles in *mir-34* mutants is significantly higher than in controls (22.2 ± 1.8 versus 1.5 ± 0.3 in medulla; 19.2 ± 2.5 versus 7.0 ± 0.9 in other regions of the brain; $**P < 0.001$, one-way analysis of variance, with post test: Tukey's multiple comparison test). Mean \pm s.e.m., $n = 10$ independent male fly brains. Genotypes as in **a**. Scale bar: 0.1 mm. **d**, *mir-34* mutant flies have a transcriptional profile indicative of accelerated ageing. Top panel: 173 age-correlated probe sets were defined from a transcriptional profile of fly brains at 3 days, 30 days and 60 days of age. Arrowheads indicate time points (3 days and 20 days) at which *mir-34* mutants and controls were compared. Genotype: iso31 flies used for transcriptional profiles of normal ageing brains. $n = 3$ biological replicates for each time point. $P \leq 0.001$, false discovery rate (FDR) ≤ 0.062 , linear regression model. Bottom panel: scatter plot illustrates the relative expression of 173 probe sets, which shows a significant difference between *mir-34* mutants and age-matched controls ($P = 0.006$, two-sample, paired Wilcoxon test). Whereas the pattern for positively correlated probe sets (red), indicated by the contour lines, is significantly different ($P = 0.0001$) between the two genotypes, and tends to show higher expression in *mir-34* mutants compared to controls, it is not for negatively correlated probe sets (blue) ($P = 0.9583$). Contour lines indicate that positively correlated probe sets tend to show higher expression in *mir-34* mutants compared to controls. Genotypes as in **a**. $n = 5$ biological replicates for each time point.

coupled with shortened lifespan. We therefore hypothesized that loss of *mir-34* accelerated brain ageing. To address this, we transcriptionally profiled the fly brain (3 days, 30 days and 60 days) from wild-type animals. On the basis of a linear regression model⁹, we extracted 173 probe sets from this profile the expression of which was tightly correlated with the progression of normal ageing (Fig. 2d and Supplementary Tables 2 and 3). We next made another set of brain transcriptional profiles for *mir-34* mutants and controls of matched chronological age (3 days and 20 days). We measured relative changes of these probe sets between 3 days and 20 days within each genotype, and compared the extent of such changes between *mir-34* mutants and controls. This indicated that the overall pattern of these probe sets was significantly different between the two genotypes ($P = 0.006$, two-sample, paired Wilcoxon test; Fig. 2d). In particular, most positively correlated probe sets displayed a faster pace of increase in *mir-34* mutants compared to controls—thus showing accelerated age-associated expression changes in *mir-34* mutants (Fig. 2d). This result, combined with the physiological and histological evidence of more rapid loss of age-associated functions, suggested that *mir-34* mutants were undergoing accelerated brain ageing.

miRNAs function by binding to the 3' UTRs of target mRNAs and often result in downregulation of protein translation. We therefore reasoned that age-associated activities of miR-34 might be mediated through silencing of critical targets that have a negative impact on the adult animal. miRNA-target prediction algorithms indicated miR-34 binding sites within the 3' UTR of the *Eip74EF* gene; notably, these binding sites were conserved in the orthologous *Eip74EF* genes from different *Drosophila* species (Supplementary Fig. 4a). We confirmed the miR-34 interaction through mutations in the seed sequences of the predicted miR-34 binding sites in the 3' UTR of the *Eip74EF* mRNA (Supplementary Fig. 4b). The *Eip74EF* gene is a component of steroid hormone signalling pathways. Although such pathways have generally been studied for effects during development, data have implicated these pathways in lifespan regulation¹⁰.

The *Eip74EF* gene encodes two major protein isoforms, E74A and E74B (referred to as the *E74A* and *E74B* genes, respectively¹¹); the isoforms share the same 3' UTR (Supplementary Fig. 4a). Northern blots indicated that transcription of *E74A*, but not *E74B*, persisted in adults, overlapping the time period when *mir-34* is expressed (Supplementary Fig. 4c). Given this, we focused on *E74A* as a regulated target of miR-34 in the adult. Despite robust expression of the mRNA transcript, the E74A protein was expressed at low levels in adult heads throughout lifespan (Fig. 3a, b and Supplementary Fig. 4d). In flies lacking miR-34, E74A protein was markedly increased (Fig. 3b); E74A was also de-regulated in the *loqs*⁰⁰⁷⁹¹ mutant flies (Supplementary Fig. 1e). Genomic rescue of *mir-34* mitigated this de-regulation of the E74A protein (Fig. 3c). Fine temporal analysis indicated that the E74A protein was highly expressed in young flies, but underwent a marked decrease within a 24-h time window (Supplementary Fig. 5). This temporal pattern seemed to be mutually exclusive to that of miR-34 (see Supplementary Fig. 2a). Moreover, in flies lacking miR-34, the downregulation of E74A protein during this critical period was dampened (Supplementary Fig. 5). This evidence indicates that adult-onset expression of *mir-34* functions, at least in part, to attenuate E74A protein expression in the young adult, and maintain that repression through adulthood (Supplementary Fig. 4d).

We next determined whether deregulated expression of E74A protein contributed to the age-associated defects in *mir-34* mutants. Because E74A function is essential during development, with strong mutations leading to pre-adult lethality¹¹, we used the mild, but viable, *E74A*^{BG01805} hypomorphic mutation (Supplementary Fig. 4a). When the *E74A*^{BG01805} mutation was combined with *mir-34* mutant flies in the same homogenous genetic background, proper regulation of E74A protein was partially restored (Fig. 3d), and age-associated defects due to loss of *mir-34*, including shortened lifespan and brain vacuolization, were mitigated (Fig. 3e, f; *E74A*^{BG01805} mutants alone have a normal lifespan

(Supplementary Fig. 6a)). To assess further the adult activity of E74A, we upregulated E74A in the adult with an *E74A* transgene that lacks miR-34 binding sites driven by a temperature-sensitive promoter¹². At 29 °C, these flies demonstrated increased levels of E74A expression in the adult (Supplementary Fig. 6b). Notably, these animals also showed late-onset brain degeneration (Supplementary Fig. 6c) and a significantly shortened lifespan (Supplementary Fig. 6d). These data indicate that deregulated expression of E74A has a negative impact on normal ageing, and that one function of miR-34 is to silence E74A in the adult to prevent the adult-stage deleterious activity of E74A on brain integrity and viability.

Notably, during the course of these studies, we noted that *mir-34* mutants also displayed a defect in protein misfolding—a molecular process implicated in ageing and common to many human neurodegenerative diseases¹³. Whereas normally with age, the fly brain accumulates a low level of inclusions that immunostain for stress chaperones like Hsp70/Hsc70, *mir-34* mutants showed a marked increase compared to control flies of matched age (30 days) (Supplementary Fig. 7). Given that *mir-34* expression increases with age, and *mir-34* loss shows altered chaperone accumulation, we tested whether *mir-34* expression itself is upregulated by stresses like heat shock or oxidative toxins, but found no evidence to support this (data not shown). However, given that loss of *mir-34* caused an increase in protein misfolding, this raised the possibility that upregulation of *mir-34* expression might mitigate disease-associated protein misfolding. In *Drosophila*, expression of a pathogenic ataxin-3 polyglutamine (polyQ) disease protein (SCA3trQ78) leads to inclusion formation, a decrease in polyQ protein solubility and progressive neural loss¹⁴ (Supplementary Fig. 8a). Upregulation of *mir-34* markedly mitigated polyQ degeneration, such that inclusion formation was slowed, the protein retained greater solubility, and neural degeneration was suppressed (Fig. 3g, h and Supplementary Fig. 8b–d). Lowering E74A expression by heterozygous reduction in flies expressing pathogenic polyQ protein revealed a minimal effect (data not shown), indicating that E74A may not be a target of miR-34 activity in this process. However, our studies with E74A were of necessity limited to hypomorphic alleles that may not uncover the full extent of E74A function mediated by miR-34. Furthermore, additional targets of miR-34 may be involved in different aspects of miR-34-directed pathways, including disease.

Given this effect to mitigate disease-associated neural toxicity with upregulation of *mir-34*, and that *mir-34* expression naturally increases with age, we investigated whether enhanced expression of *mir-34* in wild-type flies could modulate the ageing process. We increased miR-34 dosage in wild-type flies with genomic rescue transgenes, which express *mir-34* under its endogenous regulatory elements (see Supplementary Fig. 2a). Analysis of multiple independent transgenics in the same genetic background with that of control indicated that upregulation of miR-34 levels with genomic constructs (~20%, Supplementary Fig. 3d) promoted median survival rate by ~10% compared to wild type (Fig. 3i; other traits, such as the occurrence of brain vacuolization, despite being an age-associated phenomenon, are sporadic and low in normal flies, thus were difficult to assess). Thus, upregulation of *mir-34* expression can protect from neurodegenerative disease and extend median lifespan.

Our findings indicate that miR-34 in *Drosophila* presents a key miRNA that couples long-term maintenance of the brain with healthy ageing of the organism. miR-34 activity, enhanced by its age-modulated expression and processing, is critically involved in silencing of the *E74A* transcript through adulthood and in modulation of protein homeostasis with age, as well as in polyQ disease. Select neural cell types may be especially vulnerable in ageing and disease¹⁵; miR-34 function may have an impact on the integrity or activity of these systems. Intriguingly, E74A seems to confer sharply opposing function on animal fitness at different life stages, being essential during pre-adult development¹¹, but harmful to the adult during ageing (this study). This biological property—of a gene being beneficial at one

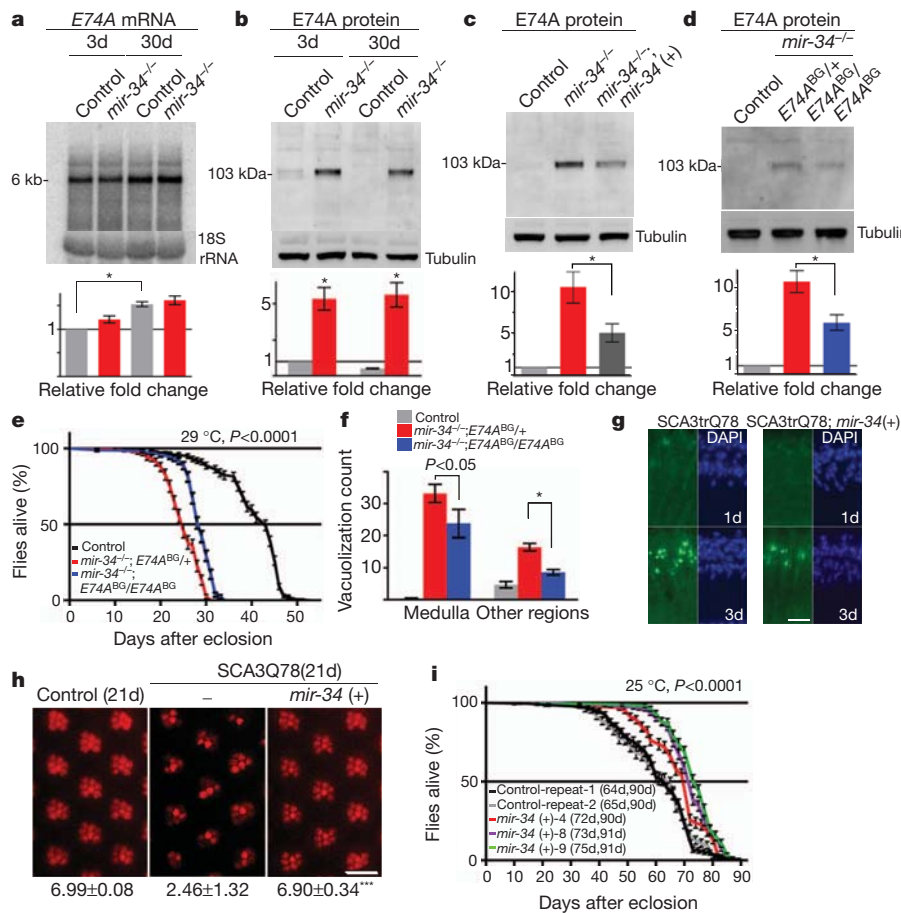


Figure 3 | The *Drosophila Eip74EF* gene is a target of miR-34 in modulation of the ageing process. **a**, *E74A* mRNA is robustly expressed in the adult and unchanged between age-matched controls and *mir-34* mutants. In control flies, *E74A* mRNA is significantly upregulated in 30 day compared to 3 day animals. RNA was from male heads. Mean \pm s.e.m., $n = 3$ independent experiments; signal density of *E74A* mRNA normalized to 18S rRNA loading control ($*P < 0.01$, one-way analysis of variance, with post test: Tukey's multiple comparison test). Genotypes: control, 5905; *mir-34*^{-/-}, *mir-34* null-1 in 5905 homogenous genetic background. **b**, *E74A* protein is deregulated in *mir-34* mutants. Protein was from male heads. Mean \pm s.e.m., $n = 3$ independent experiments; signal density of *E74A* protein normalized to tubulin loading control ($*P < 0.01$, one-way analysis of variance, with post test: Tukey's multiple comparison). Genotypes as in **a**. **c**, Deregulation of *E74A* protein is diminished in *mir-34* rescue flies. Protein was from male heads. Mean \pm s.e.m., $n = 3$ independent experiments; signal density normalized to tubulin loading control ($*P < 0.05$, one-way analysis of variance, with post test: Tukey's multiple comparison test). Genotypes: control, 5905; *mir-34*^{-/-}, *mir-34* null-1 in 5905 homogenous genetic background; *mir-34*^{-/-}; *mir-34*(+), *mir-34* genomic rescue in *mir-34* null-1 in 5905 homogenous genetic background. **d**, *mir-34* mutants homozygous for the *E74A*^{BG01805} allele have lower levels of *E74A* protein. Protein was from male heads of 20 day flies raised at 29 °C. Mean \pm s.e.m., $n = 3$ independent experiments; signal density normalized to tubulin loading control ($*P < 0.01$, one-way analysis of variance, with post test: Tukey's multiple comparison test). Genotypes: control, 5905; *mir-34*^{-/-} *E74A*^{BG}/+, *E74A*^{BG01805}/+, *mir-34* null-1 in 5905 homogenous genetic background; *mir-34*^{-/-} *E74A*^{BG}/*E74A*^{BG}, *E74A*^{BG01805}/*E74A*^{BG01805}, *mir-34* null-1 in 5905 homogenous genetic background. **e**, **f**, Reducing *E74A* protein levels in the adult mitigates age-related defects of *mir-34* mutants. *mir-34* mutants also homozygous for *E74A*^{BG01805} show rescued lifespan (**e**) and brain morphology (**f**), compared to *mir-34* mutants heterozygous for *E74A*^{BG01805} (these flies have a lifespan that is the same as *mir-34* mutants alone; see

Supplementary Table 4). Flies raised at 29 °C. Lifespan: $P < 0.0001$ (log-rank test). Mean \pm s.e.m., $n \geq 150$ male flies. Brain vacuoles: $*P < 0.01$ (one-way analysis of variance, with post test: Tukey's multiple comparison test). Mean \pm s.e.m., $n = 10$ independent male animals. Genotypes as in **d, g**. Upregulation of *mir-34* reduces accumulation of pathogenic polyQ protein inclusions. Left panels: in the retina of flies expressing SCA3trQ78 alone, pathogenic polyQ protein is initially diffuse (1 day, top), but gradually accumulates into nuclear inclusions (3 day, bottom). Right panels: upregulation of *mir-34* reduces inclusion formation. DAPI staining highlights nuclei. 3 day controls show 53.75 ± 12.55 inclusions in a retinal section versus 23.67 ± 7.57 with *mir-34* upregulation; mean \pm s.d., $n \geq 3$ cryosections from independent male animals; $P < 0.01$ (*t*-test). Genotypes: SCA3trQ78 is w^+ ; *rh1-GAL4, UAS-SCA3trQ78/+*. SCA3trQ78; *mir-34 (+)* is w^+ ; *rh1-GAL4, SCA3trQ78/+; UAS-mir-34/+*. Scale bar, 0.05 mm. **h**, Upregulation of *mir-34* prevents neural degeneration. At 21 days, male flies expressing SCA3trQ78 show a marked loss of photoreceptor neuronal integrity (middle panel), with an average of only 2.46 ± 1.32 photoreceptors per ommatidium remaining by pseudopupil analysis. Flies with upregulated *mir-34* (right panel) retain 6.90 ± 0.34 photoreceptors per ommatidium. Control (left panel) and upregulation of *mir-34* alone (not shown) have normal photoreceptor numbers per ommatidium. Mean \pm s.d., $n = 619, 722$ and 700 ommatidia, for SCA3trQ78, SCA3trQ78; *mir-34 (+)* and control, respectively; $***P < 0.0001$ (one-way analysis of variance, with Bonferroni's multiple comparison test). Genotypes as in **b**; control: w^+ ; *rh1-GAL4/+*. Scale bar, 0.05 mm. **i**, Flies with upregulated *mir-34* (colour) have an extended median lifespan compared to control flies (black and grey curves for repeats 1 and 2, respectively) (log-rank test). Lifespan result for each genotype is indicated in median and maximal days. Mean \pm s.e.m., $n \geq 150$ male flies per genotype, 25 °C. Three independent *mir-34* genomic transgenic lines (4, 8, 9) were analysed. Genotypes: control, 5905; *mir-34 (+)*, *mir-34* genomic rescue in 5905 homogenous genetic background.

life stage, but damaging at another—is referred to as antagonistic pleiotropy¹⁶. Genes associated with antagonistic pleiotropy are likely to be evolutionarily retained due to their earlier beneficial function¹⁷. Their adult-onset activities, however, antagonize the ageing process if

they are not properly regulated. miRNA pathways provide a tantalizing mechanism by which to suppress potentially deleterious age-related activities of such genes; a number of miRNAs have been noted to show age-modulated expression and activity^{18,19}. Roles of

select miRNAs normally expressed in the adult may be of evolutionary advantage to tune-down events that promote age-associated decline and potentially disease, in order to prolong healthy lifespan and longevity. Upregulation of *lin-4*, a *C. elegans* miRNA with a known developmental role, extends nematode lifespan¹⁸, raising the possibility that this upregulation, like the natural increase of *mir-34* expression in *Drosophila*, functions to silence genes that have a negative impact on ageing and potentially promote disease. Notably, *mir-34* expression is elevated with age in *C. elegans*^{19,20}, and mammalian *mir-34* orthologues are highly expressed in the adult brain²¹ and have also been noted to increase with age and be misregulated in degenerative disease in humans^{22–26}. Current data regarding miR-34 function indicate that it is neutral or adverse in *C. elegans*^{19,27}, and can be either protective or contributory to age-associated events in vertebrates^{22–26}. Thus, miR-34 seems to be a key miRNA poised to integrate age-associated physiology; the precise function will reflect the diverse spatiotemporal expression and activity of distinct orthologues, the mRNA target spectrum, as well as the complexity of the adult brain and life cycle. The conservation of miR-34, coupled with in-depth comparative analysis of *mir-34* expression, 3' end processing, targets and pathways in the ageing process of nematodes, flies and mammals, make it a tempting subject for understanding features of ageing and disease susceptibility.

METHODS SUMMARY

Flies were grown in standard media at 25 °C unless otherwise specified. Stock lines and *GAL4* driver lines were obtained from the *Drosophila* Stock centre at Bloomington, or are described¹⁴. Deletion of the *mir-34* region was made by site-specific recombination. Fly transgenics were generated by standard procedures. Flies were generated or backcrossed a minimum of five generations into a controlled uniform homogeneous genetic background (line 5905 (FlyBase ID FBst0005905, *w¹¹¹⁸*)), to assure that all phenotypes were robust and not associated with variation in genetic background. In this uniform homogeneous genetic background, the lifespan of control flies is highly uniform with repetition when 150 or more individuals are used for lifespan analysis. Negative geotaxis and thermo stress were used to examine fly locomotion and stress resistance, respectively. Adult male heads were processed for paraffin sections as described¹⁴. To determine lifespan, newly eclosed males were collected and maintained at 15 flies per vial, transferred to fresh vials every 2 days while scored for survival. A total of 150–200 flies were used per genotype per lifespan; all experiments were repeated multiple times (see Supplementary Table 4). Lifespans were analysed in Excel (Microsoft) and by Prism software (GraphPad) for survival curves and statistics. Techniques of molecular biology, western immunoblots and histology were standard. Fly brain mRNA was prepared using Trizol reagent for array and mRNA analysis, miRNA arrays were miRCURY LNA arrays version 8.1 (Exiqon), and mRNA expression was profiled using Affymetrix *Drosophila* 2.0 chips (Affymetrix). The microarray data can be found in the Gene Expression Omnibus (GEO) of NCBI through accession number GSE25009.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 11 June; accepted 22 December 2011.

Published online 15 February 2012.

1. Amaducci, L. & Tesco, G. Aging as a major risk for degenerative diseases of the central nervous system. *Curr. Opin. Neurol.* **7**, 283–286 (1994).
2. Eacker, S. M., Dawson, T. M. & Dawson, V. L. Understanding microRNAs in neurodegeneration. *Nature Rev. Neurosci.* **10**, 837–841 (2009).
3. Bilen, J., Liu, N., Burnett, B. G., Pittman, R. N. & Bonini, N. M. MicroRNA pathways modulate polyglutamine-induced neurodegeneration. *Mol. Cell* **24**, 157–163 (2006).
4. Jiang, F. *et al.* Dicer-1 and R3D1-L catalyze microRNA maturation in *Drosophila*. *Genes Dev.* **19**, 1674–1679 (2005).

5. Liu, N. *et al.* The exoribonuclease Nibbler controls 3' end processing of microRNAs in *Drosophila*. *Curr. Biol.* **21**, 1888–1893 (2011).
6. Han, B. W. H. *et al.* n. g. J. H., Weng, Z., Zamore, P. D. & Ameres, S. L. The 3'-to-5' exoribonuclease Nibbler shapes the 3' ends of microRNAs bound to *Drosophila* Argonaute1. *Curr. Biol.* **21**, 1878–1887 (2011).
7. Chung, W. J., Okamura, K., Martin, R. & Lai, E. C. Endogenous RNA interference provides a somatic defense against *Drosophila* transposons. *Curr. Biol.* **18**, 795–802 (2008).
8. Kretschmar, D., Hasan, G., Sharma, S., Heisenberg, M. & Benzer, S. The *swiss cheese* mutant causes glial hyperwrapping and brain degeneration in *Drosophila*. *J. Neurosci.* **17**, 7425–7432 (1997).
9. Cao, K., Chen-Plotkin, A. S., Plotkin, J. B. & Wang, L. S. Age-correlated gene expression in normal and neurodegenerative human brain tissues. *PLoS ONE* **5** (2010).
10. Simon, A. F., Shih, C., Mack, A. & Benzer, S. Steroid control of longevity in *Drosophila melanogaster*. *Science* **299**, 1407–1410 (2003).
11. Fletcher, J. C. & Thummel, C. S. The *Drosophila* E74 gene is required for the proper stage- and tissue-specific transcription of ecdysone-regulated genes at the onset of metamorphosis. *Development* **121**, 1411–1421 (1995).
12. Fletcher, J. C., D'Avino, P. P. & Thummel, C. S. A steroid-triggered switch in E74 transcription factor isoforms regulates the timing of secondary-response gene expression. *Proc. Natl Acad. Sci. USA* **94**, 4582–4586 (1997).
13. Morimoto, R. I. Proteotoxic stress and inducible chaperone networks in neurodegenerative disease and aging. *Genes Dev.* **22**, 1427–1438 (2008).
14. Warrick, J. M. *et al.* Expanded polyglutamine protein forms nuclear inclusions and causes neural degeneration in *Drosophila*. *Cell* **93**, 939–949 (1998).
15. Hirth, F. *Drosophila melanogaster* in the study of human neurodegeneration. *CNS Neurol. Disord. Drug Targets* **9**, 504–523 (2010).
16. Williams, G. C. Pleiotropy, natural selection and the evolution of senescence. *Evolution* **11**, 398–411 (1957).
17. Kirkwood, T. B. Understanding the odd science of aging. *Cell* **120**, 437–447 (2005).
18. Boehm, M. & Slack, F. A developmental timing microRNA and its target regulate life span in *C. elegans*. *Science* **310**, 1954–1957 (2005).
19. de Lencastre, A. *et al.* MicroRNAs both promote and antagonize longevity in *C. elegans*. *Curr. Biol.* **20**, 2159–2168 (2010).
20. Ibanez-Ventoso, C. *et al.* Modulated microRNA expression during adult lifespan in *Caenorhabditis elegans*. *Aging Cell* **5**, 235–246 (2006).
21. Bak, M. *et al.* MicroRNA expression in the adult mouse central nervous system. *RNA* **14**, 432–444 (2008).
22. Zoviolis, A. *et al.* microRNA-34c is a novel target to treat dementias. *EMBO J.* **30**, 4299–4308 (2011).
23. Minones-Moyano, E. *et al.* MicroRNA profiling of Parkinson's disease brains identifies early downregulation of miR-34b/c which modulate mitochondrial function. *Hum. Mol. Genet.* **20**, 3067–3078 (2011).
24. Li, X., Khanna, A., Li, N. & Wang, E. Circulatory miR34a as an RNA based, noninvasive biomarker for brain aging. *Aging* **3**, 985–1002 (2011).
25. Khanna, A., Muthusamy, S., Liang, R., Sarojini, H. & Wang, E. Gain of survival signaling by down-regulation of three key miRNAs in brain of calorie-restricted mice. *Aging* **3**, 223–236 (2011).
26. Gaughwin, P. M. *et al.* Hsa-miR-34b is a plasma-stable microRNA that is elevated in pre-manifest Huntington's disease. *Hum. Mol. Genet.* **20**, 2225–2237 (2011).
27. Yang, J. *et al.* MiR-34 modulates *Caenorhabditis elegans* lifespan via repressing the autophagy gene *atg9*. Age. doi:10.1007/s11357-011-9324-3 (2011).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank C. Thummel, T. Jongens and A. Bashirullah for reagents. We are grateful to A. Cashmore, A. Burguete, J. Kim, S. Cherry, B. Gregory, A. Gitler and the Bonini laboratory for discussion and critical reading of the manuscript. We thank X. Teng for assistance with fly paraffin section. This work was funded by the NINDS (R01-NS043578) and the Ellison Foundation (to N.M.B.). L.-S.W. and K.C. are supported by a pilot grant from Penn Genome Frontiers Institute. L.-S.W. is supported by NIA (U01-AG-032984-02 and RC2-AG036528-01) and a Penn Institute on Aging pilot grant (AG010124). N.M.B. is an Investigator of the Howard Hughes Medical Institute. J.R.K. received support from NIH T32 AG00255.

Author Contributions N.L. and N.M.B. conceived and designed the project. N.L., M.L., M.A., G.-J.H., J.R.K. and Y.Z. planned, executed and analysed experiments. K.C. and L.-S.W. performed aging computational modelling. N.L. and N.M.B. wrote the manuscript with input from all authors.

Author Information The microarray data can be found in the Gene Expression Omnibus (GEO) of NCBI through accession number GSE25009. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to N.M.B. (nbonini@sas.upenn.edu).

METHODS

Genetic background. Fly lines were from the Bloomington Stock centre or are described¹⁴. To control for background effects, and to assess significance of all effects, flies were generated in the same uniform homogeneous genetic background (line 5905 (FlyBase ID FBst0005905, *w¹¹¹⁸*)), or backcrossed a minimum of five generations into this uniform genetic background. This assured that, for all phenotypes, even modest and consistent effects were associated with the gene manipulations and not a variation in background. With these carefully controlled experiments, the lifespan of control flies was highly uniform upon repetition, when 150 or more individuals were used for lifespan analysis (see Supplementary Table 4).

mir-34 deletion mutants. Deletion of the *mir-34* region was made by site-specific recombination between two piggyBac insertions, using FLP-FRT-mediated site-specific recombination²⁸. The loss of other genes in the region was then fully rescued by genomic transgenes, so that a line selectively lacking only *mir-34* was generated. Two FRT-bearing insertions, PBac[XP]d02752 and PBac[RB]Fmr1⁶⁰²⁷⁹⁰, were used (Exelixis collection), which encompass the *mir-34* region. Genetic crosses were made to combine these two transposon elements with heat-inducible FLP recombinase. After 48 h of egg laying, parents were removed, and vials containing progeny were placed in a 37 °C water bath for a 1-h heat shock. Progeny flies were treated with daily 1-h heat shock, for an additional 4 days. Young virgin female progeny flies were collected and crossed to males with 3rd chromosome balancers. In the subsequent generation, progeny males were used to generate additional progeny for PCR confirmation. Progeny flies bearing the deletion were positive for PCR verification, using primers from neighbouring genomic DNA and ones from transposons (upstream insertion: 5'-GGTCGTGCATGACGAGATTA-3'/5'-TACTATTCTTCTCACTCGCAGTTATTG-3'; downstream insertion: 5'-TC CAAGCGCGACTGAGATG-3'/5'-GTGCGTTCGAAGAAATGATG-3'). Flies with the *mir-34* region deletion were viable, and were further verified for the appropriate deletion by PCR amplification, with primers for *mir-317* (5'-CGGAAA AACGGTTTGTGTCT-3'/5'-CCCGGGAACGAGTAAACGAAATGAAATCA-3'), *mir-277* (5'-TGATTATGGTTTGTGTTTCAGTTG-3'/5'-TTGATATCATT TCACACTATCACAAAAATGTC-3'), *mir-34* (5'-ACCTTGAGCGCTTCAAC TCT-3'/5'-CACTCTTCTCGTTTGCATGG-3') and *dfmr1* (5'-CACACAGA GCTTCCCACTGA-3'/5'-AGGCCCTCTTTTGTGACATT-3').

Fly age-associated phenotypes. Negative geotaxis and thermo stress were used to examine fly locomotion and stress resistance, respectively. To perform negative geotaxis, groups of 15 adult male flies of indicated age were transferred into a 14-ml polystyrene round-bottom tube (Falcon), and placed in the dark for 30-min recovery. The assay was conducted in the dark, with only a red light on. Climbing ability was scored as the percentage of flies failing to climb higher than 1.5 cm from the bottom of the tube, within 15 s after gently being banged to the bottom. Three repeats were performed for each group and the result averaged. For each genotype at a given age, a minimum of 200 flies were tested. For heat sensitivity, groups of 15 adult males of indicated age were transferred into 14-ml polystyrene round-bottom tubes (Falcon) then placed in a 25 °C incubator for 30 min recovery. Heat stress was applied by immersing the vial containing the flies into a 37 °C water bath for 1 h, followed by a 30-min recovery at 25 °C, then another 1-h heat stress at 37 °C. Flies were then transferred into regular food vials and maintained at 25 °C. Dead flies were counted after 24 h. To assess brain morphology, adult male heads were processed for paraffin sections as described²⁹, and brain vacuoles were counted through continuous sections generated from each head (*n* = 10 heads counted for each genotype).

Molecular biology. Fly genomic DNA was prepared from whole flies with the Puregene DNA purification kit (Qiagen). To generate *mir-34* pUAST constructs, PCR amplification was conducted using genomic DNA as template, with primer pairs of pUAST *mir-34*-I (286 bp, PCR primer 5'-CCGTTACACAGCACT ATTCTCAAT-3'/5'-CCATCTGATACAGGTCCTACATTTTCTAAAA-3') and pUAST *mir-34*-II (936 bp, PCR primer 5'-ACCTTGAGCGCTTCAACTCT-3'/5'-CACTCTTCTCGTTTGCATGG-3'). PCR products were then ligated into the pUAST vector. *mir-277/dfmr1* rescue construct was made in the pCaSpeR4 vector, which contained two parts. Part 1 was a genomic DNA fragment (7,530 bp) harbouring the *mir-277* sequence (PCR primers: 5'-GGTCGTGCATGACGAG ATTA-3'/5'-GGATGTTTTGCGACCACTT-3'), and part 2 was a genomic fragment containing *dfmr1* genomic sequence, derived from the pBS WTR construct (a gift from T. Jongens³⁰), by BamHI and PpuMI. The *mir-34* genomic rescue construct was also made in the pCaSpeR4 vector, with two parts. One was a genomic DNA fragment (6,855 bp) upstream of *mir-277* sequence (PCR primers: 5'-GGTCGTGCATGACGAGATTA-3'/5'-GGATGCATTTTATCGTT AGGC-3'), and the other was a genomic DNA fragment (2,111 bp) containing *mir-34* sequence (PCR primers: 5'-GCAGGAAAATGCGATAAATGA-3'/5'-TCGTTACAACATGGAATCCTC-3'). The resultant construct, therefore, contains *mir-34* sequence, including most upstream fragment, with the exclusion

of 108 bp of *mir-277* sequence. In addition, a modified *mir-34* genomic rescue construct was made (pCaSpeR4 vector), which contains same upstream and downstream ends of the original *mir-34* genomic rescue construct, with a small deletion of *mir-277* mature sequence. The genomic regulation of *mir-34* seems complex, as despite these standard manipulations for gene rescue, the genomic rescue expression of *mir-34* and extent of phenotypic rescue of *mir-34* mutants was only partial. We attempted upregulation of *mir-34* with the GAL4-UAS system, including with the conditional gene switch system in adults. Upregulation of *mir-34* during development in non-germline tissues (when it normally is not expressed; Supplementary Fig. 2a) was deleterious, and we were unable to upregulate *mir-34* expression more robustly than with the genomic constructs.

For western immunoblots, 10 adult male heads per sample were homogenized in 50 µl of Laemmli buffer (Bio-Rad) supplemented with 5% 2-mercaptoethanol, heated to 95 °C for 5 min and 10 µl loaded onto 4–12% Bis-Tris gels (NuPage), then transferred to nitrocellulose membrane (Biorad) and blotted by standard protocols. Primary antibodies used were anti-tubulin (1:10,000, E7, Developmental Studies Hybridoma Bank), anti-E74A (a gift of C. Thummel). Secondary antibodies for immunoblots were goat anti-mouse conjugated to HRP (1:2,000, Chemicon) and developed by chemiluminescence (ECL, Amersham). The final image was obtained by Fuji scanner (Fujifilm).

Total RNA was isolated from 50–200 male heads per genotype, by cutting off heads with a sharp razor, then putting heads into Trizol reagent. Heads were ground by pestle, then RNA was isolated following the manufacturer's protocol (Trizol reagent, Invitrogen). 5 µg RNA was used per lane. Gel running (1% agarose) and blot transfer (nylon plus) were according to recommended procedures (Northernmax, Ambion). The RNA blot was then used for hybridization following standard procedures at 68 °C, with pre-hybridization (~1 h), hybridization (~12 h or overnight) with P³²-labelled probe, washed and exposed to Phosphorimager (Amersham). RNA probes were used that were made by *in vitro* transcription of cDNA templates using Maxiscript-T7 *in vitro* transcription kit (Ambion), supplemented with P³²-labelled UTP. The cDNA templates were prepared from total RNA by one-step RT-PCR (SuperScript One-Step RT-PCR with Platinum Taq, Invitrogen), with primers: *E74A* (5'-GTGAACGTGGTGGTGAAC-3'/5'-GATAATACGACTCACTATAGGGAGATGTCCATTGCTTCTCAATG-3'); *E74B* (5'-CATCGCTTGTCAATGTGTCC-3'/5'-GATAATACGACTCACTA TAGGGAGACTGCGGTAATCACTGAGCTG-3'); 18S rRNA loading control (5'-GATAATACGACTCACTATAGGGAGA-3'/5'-AGGGAGCCTGAGAAAC GGCTACCACATCTAAGGAATCTCCCTATAGTGAGTCGTATTATC-3').

For small RNA northern blots, total RNA was isolated from male fly heads using Trizol reagent as above. For each lane, 3 µg of RNA was used, and RNA was fractionated on a 15% Tris-UREA gel (NuPage) with 1×TBE buffer. The blot transfer was performed with 0.5×TBE buffer. Before hybridization, the RNA blots were pre-hybridized with Oligohyb (Ambion), and then incubated with radioactive labelled RNA probes for ~12 h to overnight. RNA probes were used, and made by *in vitro* transcription of oligo templates using Maxiscript-T7 *in vitro* transcription kit (Ambion), supplemented with P³²-labelled UTP. Oligo DNA templates were prepared by annealing two single-stranded DNA oligonucleotides into duplex (99 °C, 5 min and cool down to room temperature). Oligonucleotides used for *mir-34* (5'-GATAATACGACTCACTATAGGGAGA-3'/5'-AAAAATGGCA GTGTGGTTAGCTGGTTGTGTCTCCCTATAGTGAGTCGTATTATC-3'), *mir-277* (5'-GATAATACGACTCACTATAGGGAGA-3'/5'-TAATGCACATCTG GTACACATAAATGCACATCTGGTACGACA TCTCCCTATAGTGAGT CGTATTATC-3') and 2S rRNA (5'-GATAATACGACTCACTATAGGGA GA-3'/5'-TGCTTGACTACATATGGTTGAGGGTTGTATCTCCCTATAGT GAGTCGTATTATC-3').

Luciferase assays were performed using standard approaches⁵. Specifically, 8 × 10⁴ DL1 cells were plated and bathed in 30 µl of serum-free medium with 60 ng of dsRNA in each well of a 96-well plate. The next day, 1.6 ng of pMT-Firefly, 400 ng of pMT-*mir-34* and 400 ng of pMT-renilla *E74A* wild-type or mutant 3' UTR reporters were transfected by Effectene (Qiagen). Two days after transfection, the expression of the reporters and *mir-34* was induced by CuSO₄. Twenty-four hours after induction, luminescence assays were performed by the Dual-Glo Luciferase Assay System (Promega). The *mir-34* seed sequences in the 3' UTR of *E74A* were mutated as noted in Supplementary Fig. 3, using the Quik change mutagenesis system (Stratagene). Primers to knockdown *vgo1* are described⁵.

The miRNA-target prediction algorithms TargetScan (v5.1)³¹ and PicTar (fly)³² were used to determine miR-34 target mRNA candidates.

miRNA microarray analysis. For miRNA array analysis, Iso31 flies (isogenized *w¹¹¹⁸*) were used. Flies were killed by brief submersion in ethanol under CO₂ anaesthesia, followed by two PBS washes (Sigma). To control for circadian effects, all flies were processed between 11:00 and 13:00. Brains were removed manually and collected in an Eppendorf microcentrifuge tube stored on ice. For each miRNA microarray replicate, 200–300 brains were collected for each time point,

with a ~50/50 ratio of males and females. RNA was prepared using the miRvana RNA extraction system (Ambion) yielding ~2.5 µg per 100 brains. RNA was eluted into 80 µl of RNase free water (Fisher Scientific) and stored at -80 °C. miRNA profiling was carried out at the Penn microarray core facility using miRCURY LNA arrays (Exiqon) and protocols. Exiqon's Hy3/H5-labelling kit was used (Exiqon). RNA samples were labelled with Hy3 and hybridized together with a Hy5-labelled common reference standard. The common reference standard consisted of equal amounts of RNA from brains of 3 days, 30 days and 60 days flies. The miRNA microarray data were analysed at the Penn Bioinformatics Core. Raw data was imported into Gene Spring 1.0 (Agilent) and normalized using a global LOESS regression algorithm (locally weighted scatterplot smoothing). Relative expression levels were calculated as the log₂ normalized signal intensity difference between the Hy3 and Hy5 intensity. Present/absent flagging was analysed by Exiqon (Exiqon). Expression levels (fold changes) for the 30 day and 60 day time point were calculated relative to the 3 day time point. The data sets were exported into Spotfire DecisionSite 9.0 (Tibco) for visualization and filtering.

mRNA microarray analysis. For ageing microarray analysis, fly stock Iso31 was used. For *mir-34* mutant microarray analysis, *mir-34* null line-1 in 5905 background was used, with fly 5905 line, as control. To generate an ageing profile, flies were aged to 3 days, 30 days and 60 days, and 30–50 brains dissected per time point, per replicate, as above (50–50 males and females). For each time point, three replicates were conducted. For *mir-34* mutant microarray analysis, time points were 3 days and 20 days, and for each time point, 20 brains from male flies of the appropriate genotype were used, with five replicates in total. Microarray hybridization and reading was performed at the Penn Microarray Core Facility. For mRNA microarrays, total RNA was reverse transcribed to ss-cDNA, followed by two PCR cycles using the Ovation RNA amplification system V2 (Ovation). Quality control on both RNA and ss-cDNA was performed using an 2100 Agilent Bioanalyzer (Quantum Analytics). The cDNA was labelled using the FL-Ovation cDNA Biotin Module V2 (Ovation), hybridized to Affymetrix *Drosophila* 2.0 chips (Affymetrix) and scanned with an Axon Instruments 4000B Scanner using GenePix Pro 6.0 image acquisition software (Molecular Devices). Affymetrix .cel (probe intensity) files were exported from GeneChip Operating Software (Affymetrix). The .cel files were imported to ArrayAssist Lite (Agilent) in which GCRMA probe-set expression levels and Affymetrix absent/present/marginal flags were calculated. Statistical analysis for those genes passing the flag filter was performed using Partek Genomics Suite (Partek). The signal values were log₂ transformed and a 2-way ANOVA was performed.

Transcriptional analysis of ageing status. We first used the wild type to extract age-associated probe sets and then compared the relative changes of these probe sets in a separate set of transcriptional profiles generated for the wild type and *mir-34* mutant. For transcriptional profiles of normal aged brains, the GCRMA package RMA (J. Z. Wu, J. MacDonald and J. Gentry, GCRMA: background adjustment using sequence information, R package version 2.14) for R/Bioconductor³³ was used to generate log₂ expression levels for probe-set IDs from the original .cel files. Then, a linear regression model was used to compute the significance of a correlation between age and gene expression⁹. This approach assumes a linear relationship between age and log₂ expression level:

$$Y_{ij} = \mu_i + \beta_{1i}A_j + \epsilon_{ij}$$

In this equation, Y_{ij} is the log₂ gene expression level of probe set i in sample j , A_j is the age for individual j . The coefficients β_{1i} is regression coefficients reflecting the rate of change in gene expression with respect to age. Probe sets with expression significantly correlated with age ($P \leq 0.001$ for β_{1i}) were determined. Then the same probe sets were used to estimate the relative expression in separate profiles of *mir-34* mutants and age-matched controls. The average levels of each individual probe set were calculated for the difference between 20 day and 3 day, within the same genotype (that is, $\Delta 20$ day/ $\Delta 3$ day) for each gene in controls and *mir-34* mutants, respectively. These differences were then compared between genotypes (that is, *mir-34* mutants – controls). The significance of the difference between genotypes was analysed using a paired Wilcoxon test. The difference between control and mutant samples in positively correlated genes (Fig. 2d) is not by chance ($P = 0.0001$).

28. Parks, A. L. *et al.* Systematic generation of high-resolution deletion coverage of the *Drosophila melanogaster* genome. *Nature Genet.* **36**, 288–292 (2004).
29. Li, L. B., Yu, Z., Teng, X. & Bonini, N. M. RNA toxicity is a component of ataxin-3 degeneration in *Drosophila*. *Nature* **453**, 1107–1111 (2008).
30. Dockendorff, T. C. *et al.* *Drosophila* lacking dfmr1 activity show defects in circadian output and fail to maintain courtship interest. *Neuron* **34**, 973–984 (2002).
31. Lewis, B. P., Shih, I. H., Jones-Rhoades, M. W., Bartel, D. P. & Burge, C. B. Prediction of mammalian microRNA targets. *Cell* **115**, 787–798 (2003).
32. Grün, D., Wang, Y. L., Langenberger, D., Gunsalus, K. C. & Rajewsky, N. microRNA target predictions across seven *Drosophila* species and comparison to mammalian targets. *PLOS Comput. Biol.* **1**, e13 (2005).
33. Gentleman, R. C. *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* **5**, R80 (2004).

Structural basis for iron piracy by pathogenic *Neisseria*

Nicholas Noinaj¹, Nicole C. Easley¹, Muse Oke¹, Naoko Mizuno², James Gumbart³, Evzen Boura¹, Ashley N. Steere⁴, Olga Zak⁵, Philip Aisen⁵, Emad Tajkhorshid⁶, Robert W. Evans⁷, Andrew R. Gorringer⁸, Anne B. Mason⁴, Alasdair C. Steven² & Susan K. Buchanan¹

***Neisseria* are obligate human pathogens causing bacterial meningitis, septicaemia and gonorrhoea. *Neisseria* require iron for survival and can extract it directly from human transferrin for transport across the outer membrane. The transport system consists of TbpA, an integral outer membrane protein, and TbpB, a co-receptor attached to the cell surface; both proteins are potentially important vaccine and therapeutic targets. Two key questions driving *Neisseria* research are how human transferrin is specifically targeted, and how the bacteria liberate iron from transferrin at neutral pH. To address these questions, we solved crystal structures of the TbpA–transferrin complex and of the corresponding co-receptor TbpB. We characterized the TbpB–transferrin complex by small-angle X-ray scattering and the TbpA–TbpB–transferrin complex by electron microscopy. Our studies provide a rational basis for the specificity of TbpA for human transferrin, show how TbpA promotes iron release from transferrin, and elucidate how TbpB facilitates this process.**

Neisseria comprise a large family of Gram-negative bacteria that colonize humans. Two family members, *Neisseria gonorrhoeae* and *Neisseria meningitidis*, are pathogens that invade the urogenital tract and nasopharynx, respectively, causing gonorrhoea, meningitis and other systemic infections. Although vaccines exist for bacterial meningitis, they have significant limitations¹ and are ineffective against serogroup B *N. meningitidis*. Currently, there are no vaccines to protect against gonococcal infections. The recent emergence of antibiotic-resistant strains² adds urgency to the need to develop more effective countermeasures for both pathogens.

Neisseria require iron for survival and virulence³. Unlike most Gram-negative bacteria, *Neisseria* do not make siderophores but instead extract iron directly from serum transferrin in the human host (TF). The neisserial transport system consists of two large surface proteins: TF-binding protein A (TbpA), a 100-kDa integral outer membrane protein belonging to the family of TonB-dependent transporters⁴; and TF-binding protein B (TbpB), an ~80-kDa co-receptor attached to the outer membrane by a lipid anchor (Supplementary Fig. 1). Both proteins are found in all clinical isolates of pathogenic *Neisseria*. TbpA binds apo and iron-containing transferrin with similar affinities, whereas TbpB only associates with iron-bound TF^{5,6}. Although TbpA can extract and import iron without TbpB, the process is considerably more efficient in the presence of the co-receptor^{7,8}. TbpA and TbpB induce bactericidal antibodies in mice against *N. meningitidis*^{9,10} and *N. gonorrhoeae*¹¹, making both proteins important vaccine targets. To elucidate how TbpA and TbpB function to bind human TF selectively and extract its tightly bound iron ($K_a = 10^{23} \text{ M}^{-1}$) at physiological pH, we combined an approach consisting of X-ray crystallography, small-angle X-ray scattering and electron microscopy to determine a model of the 260-kDa iron import complex from *N. meningitidis* strain K454 (serogroup B). Because *N. gonorrhoeae* strains FA1090 and FA19 express TbpA proteins that are 94% identical to the meningococcal protein, whereas the corresponding TbpB proteins

are 61% and 69% identical, respectively, our results are relevant to both pathogens.

Crystal structure of the TbpA–(apo)hTF complex

Structural characterization of the neisserial iron import machinery was initiated by crystallizing *N. meningitidis* TbpA with full-length, glycosylated apo human transferrin (hTF) and solving the structure to a resolution of 2.6 Å (Fig. 1, Supplementary Figs 2–4, 17 and Supplementary Table 1). Despite being significantly larger (~20%) than other structurally characterized TonB-dependent transporters⁴, TbpA retains the classic fold with a 22-strand transmembrane β -barrel encompassing a plug domain (Fig. 1a). Most of the additional mass is found in several extracellular loops which extend up to ~60 Å above the outer membrane. A plug loop implicated in iron uptake^{12,13} is unusually long and protrudes ~25 Å above the cell surface.

Human TF is a bilobal glycoprotein (~80 kDa) with a single ferric (Fe^{3+}) iron tightly bound within a cleft in each lobe (Fig. 1a and Supplementary Fig. 8). Each lobe of TF consists of two subdomains which form the cleft: N1, N2, C1 and C2. In the absence of iron, each lobe adopts an open conformation (Protein Data Bank (PDB) accession code 2HAV)¹⁴. To obtain the best model of neisserial iron import, we solved the structure of diferric hTF at 2.1 Å resolution (Fig. 2d and Supplementary Fig. 8). In our diferric structure, each lobe is found in a fully closed conformation, nearly identical to the diferric structures for both porcine (PDB code 1H76) and rabbit (PDB code 1JNF) TF.

When TbpA binds hTF, it sequesters ~2,500 Å² of buried surface, with 81 TbpA residues and 67 hTF residues participating in the interaction (Fig. 1a, c, e and Supplementary Table 2). TbpA binds exclusively to the C lobe of hTF, where electrostatic complementarity exists between the extracellular surface of TbpA (electropositive) and the C1 subdomain of hTF (electronegative) (Fig. 1b, d). Two notable features of the interface include: (1) the unusually long TbpA plug loop (residues 121–139) interacts directly with the C1 subdomain

¹Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, US National Institutes of Health, Bethesda, Maryland 20892, USA. ²Laboratory of Structural Biology, National Institute of Arthritis and Musculoskeletal and Skin Diseases, US National Institutes of Health, Bethesda, Maryland 20892, USA. ³Biosciences Division, Argonne National Laboratory, Argonne, Illinois 60439, USA. ⁴Department of Biochemistry, University of Vermont, College of Medicine, 89 Beaumont Avenue, Burlington, Vermont 05405, USA. ⁵Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, New York 10461, USA. ⁶Department of Biochemistry and Beckman Institute, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA. ⁷Metalloprotein Research Group, Division of Biosciences, School of Health Sciences and Social Care, Brunel University, Uxbridge, Middlesex UB8 3PH, UK. ⁸Health Protection Agency, Porton Down, Salisbury SP2 8NY, UK.

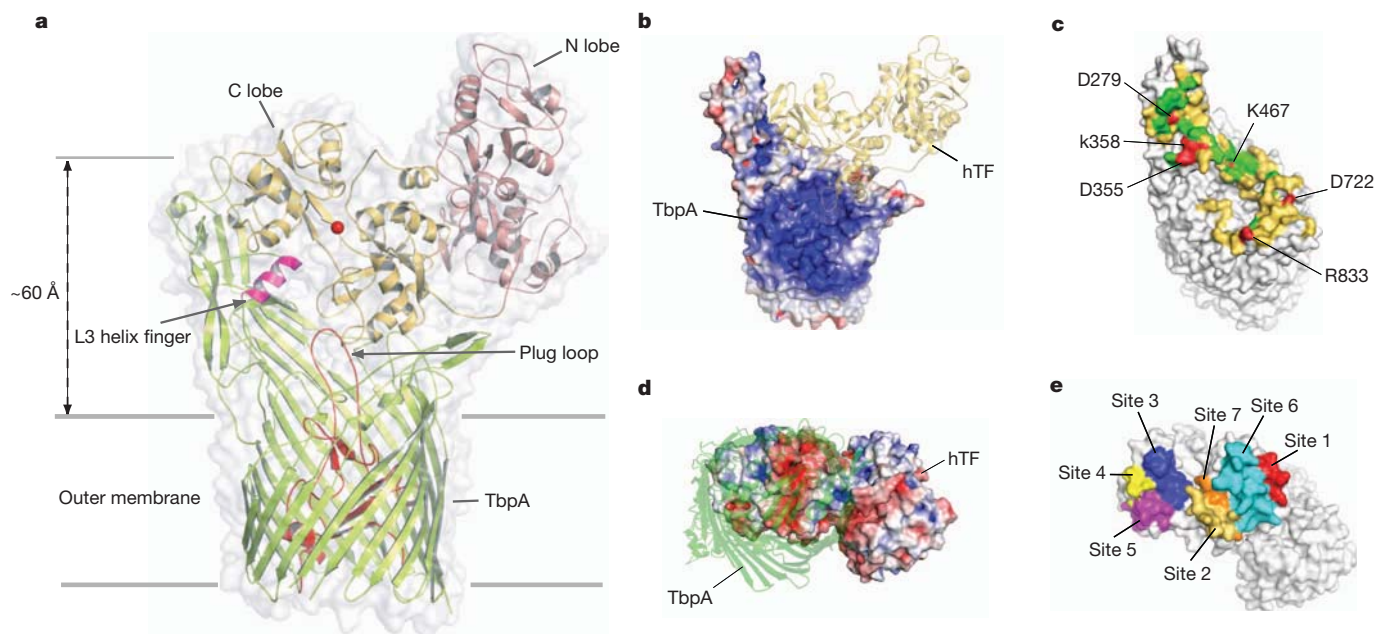


Figure 1 | Crystal structure of the TbpA–(apo)hTF complex. **a**, The TbpA β -barrel is lime, the plug is red, the helix finger is magenta. For hTF, the C lobe is gold and the N lobe is salmon. A ferric ion has been modelled into the C lobe as a red sphere. **b**, Electrostatic potential of TbpA viewed from the extracellular surface with hTF shown in gold ribbon. **c**, Residues of TbpA that bind hTF:

gold, hydrophobic interactions; green, hydrogen bonds; red, salt bridges (residues labelled). **d**, Electrostatic potential of hTF viewed from the extracellular surface with TbpA shown in green ribbon. **e**, Surface representation of hTF showing regions that bind TbpA.

(Figs 1a, 2a, c and Supplementary Fig. 5), and (2) an α -helix in TbpA extracellular loop 3 (residues 351–361, the L3 ‘helix finger’) is inserted directly into the cleft of the C lobe between the C1 and C2 subdomains (Figs 1a, 2a, b and Supplementary Fig. 5). The interaction between TbpA and hTF was found to be relatively tolerant to point mutations in TbpA, as might be expected given the large binding interface (Supplementary Fig. 6).

The mechanism of species specificity of neisserial TbpA for hTF is unknown. In *in vitro* assays, gonococcal and meningococcal TbpA proteins have been shown to bind human TF, but not TF from cow, horse, rabbit, mouse, rat, sheep, duck or pig^{15–17}. In addition, mice

infected with *N. meningitidis* displayed a higher mortality rate when the iron source was Fe_2hTF rather than bovine TF¹⁸. From the TbpA–hTF crystal structure, seven sites spanning both the C1 and C2 subdomains of hTF participate in binding TbpA (Fig. 1e and Supplementary Fig. 7), with each site containing one or more residues unique to human TF.

Because TbpA shows limited sequence variation (Supplementary Fig. 5 and Supplementary Table 4), is present in all clinical isolates, and nearly all the interactions with hTF are mediated by extracellular loops of TbpA, we attempted to disrupt the TbpA–hTF interface to see if this would be a viable therapeutic strategy. Peptides from four loops

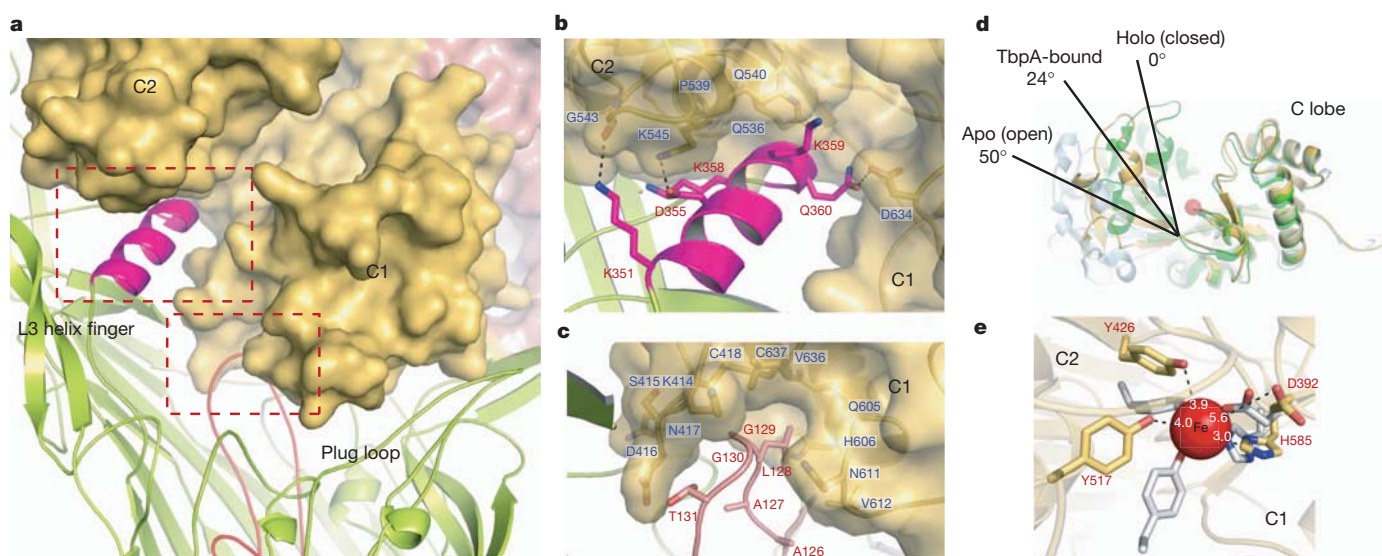


Figure 2 | TbpA distorts the iron coordination site in the hTF C lobe by inserting a helix from extracellular loop three. **a**, TbpA (green) inserts a helix finger from extracellular loop 3 (magenta) into the cleft of the hTF C lobe (gold). **b**, The helix finger interacts with the hTF C-lobe residues through main-chain and side-chain interactions. **c**, The long TbpA plug loop (pink) interacts

with residues from the C1 subdomain. **d**, Comparison of C-lobe conformations for holo (green), apo (grey, PDB code 2HAU), and TbpA-bound TF (gold). **e**, Superposition of residues coordinating iron in diferric hTF (grey) with the same residues in hTF when bound to TbpA (gold). Distances for the residues coordinating iron are shown for the TbpA-bound state.

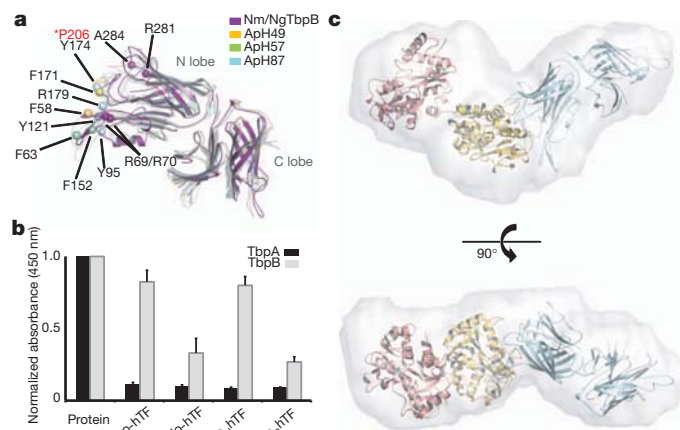


Figure 3 | SAXS analysis of the *N. meningitidis* TbpB-(holo)hTF complex. **a**, Superposition of TbpB from *N. meningitidis* with three TbpB structures from porcine pathogens. Whereas C lobes align closely, the N lobes show sequence and structural variability. Residues that diminish hTF binding when mutated are shown as spheres. Position 206 (proline in *N. meningitidis*) is critical for interaction with hTF. **b**, Competition ELISA showing relative binding affinities of TbpA and TbpB for apo-hTF, holo-hTF, hTF-Fe_N and hTF-Fe_C. Each experiment was performed in triplicate and data reported with standard errors. **c**, Fitting of the TbpB-(holo)hTF complex model into the averaged *ab initio* envelope was performed using Chimera. TbpB is shown in cyan, and hTF is shown in salmon (N lobe) and gold (C lobe).

(loops 3, 7, 11 and the plug loop) that make substantial contacts with hTF were used as antigens for polyclonal antibody development (Supplementary Table 3). All four antibodies reduced hTF binding (Supplementary Fig. 9). These results show that although the TbpA-hTF interface is extensive, reagents can be designed to disrupt it.

TbpA induces a conformational change in the hTF C lobe

In the full-length apo-hTF structure¹⁴, both the N and C lobes are in 'open' conformations, with 59.4° and 49.5° rotations required to align subdomains with diferric hTF (Fig. 2d and Supplementary Fig. 8). In the TbpA-hTF complex crystal structure, the N lobe is in the fully

open conformation. Notably, interaction with TbpA causes the C lobe to adopt a conformation midway between open and closed, with a 24° rotation required to align C1 and C2 subdomains with diferric hTF (Fig. 2d and Supplementary Fig. 10). The TbpA L3 helix finger is inserted into the cleft, where it interacts with D634 from the C1 subdomain and several residues from the C2 subdomain (Fig. 2a, b). The long TbpA plug loop also interacts with the surface of the C1 subdomain of hTF (Fig. 2c). These interactions induce a partial opening of the cleft in the hTF C lobe, thereby destabilizing the iron coordination site to facilitate the release of iron from the C lobe to TbpA. Figure 2e shows the residues coordinating iron in the hTF C-lobe structure and the significant increase in these distances when hTF binds TbpA. Such increases are clearly incompatible with tight binding of iron in the C lobe.

X-ray and SAXS structures of TbpB and TbpB-(holo)hTF

Although only TbpA can acquire iron from hTF, the reaction is enhanced by expression of the co-receptor TbpB^{7,8}, which preferentially binds holo-hTF⁶. To understand how TbpB facilitates iron extraction and uptake, we solved the structure of *N. meningitidis* TbpB (Fig. 3a, Supplementary Figs 12, 18 and Supplementary Table 1). TbpB consists of N and C lobes that are structurally similar, sharing an eight-strand β -barrel subdomain flanked by a four-strand 'handle' domain.

TbpB proteins from different isolates vary substantially in size and sequence (Supplementary Fig. 11), but the overall fold is conserved. Our structure aligns closely with three TbpB structures from porcine pathogens^{19,20} and shows that sequence and conformational variations are found primarily in the N lobe (Fig. 3a and Supplementary Fig. 12b). Specifically, residues affecting TF binding lie on the distal surface of the N lobe^{19–22} and confer much of the TF species specificity (Fig. 3a). We made point mutants at four sites on this surface that reduced or abolished binding to hTF (Fig. 3a and Supplementary Fig. 12). An analysis of our mutants, and those reported previously, demonstrates that the major site of interaction lies in the N lobe.

To clarify the interactions of purified TbpA and TbpB with hTF, an enzyme-linked immunosorbent assay (ELISA) was used to probe binding to Fe₂hTF, monoferric hTF with iron only in the N lobe (Fe_N), or in the C lobe (Fe_C), and apo-hTF (incapable of binding iron

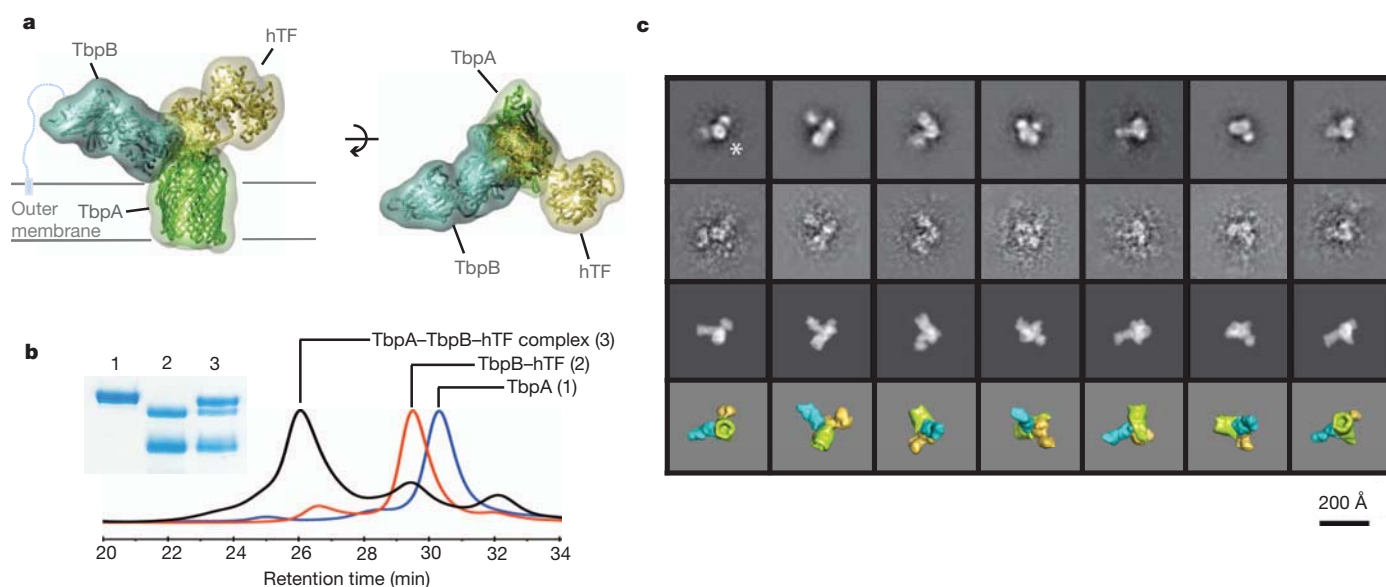


Figure 4 | Analysis of the TbpA-TbpB-(holo)hTF triple complex by negative stain electron microscopy. **a**, Model of the triple complex. TbpA is green, TbpB is light blue and hTF is gold. **b**, Purification of the triple complex by size-exclusion chromatography. The Coomassie-stained SDS gel shows purified TbpA in lane 1, the TbpB-(holo)hTF complex in lane 2, and the TbpA–

TbpB-(holo)hTF complex in lane 3. **c**, Row 1, set of seven non-redundant class averages of negatively stained complexes; row 2, examples of individual images assigned to the respective classes; row 3, re-projections of the model in the corresponding orientations, band-limited to 15 Å resolution; row 4, surface renderings of the band-limited model in the corresponding orientations.

in either lobe)²³. Consistent with earlier studies using apo- or holo-hTF⁶, TbpA binds all four hTFs with equal affinity regardless of the iron status of either lobe (Fig. 3b). In contrast, TbpB has a strong preference for hTF constructs with iron bound in the C lobe, regardless of the coordination state of the N lobe. These experiments clearly show that, at least *in vitro*, hTF interacts with TbpA and TbpB solely through the C lobe and is not affected by the presence or absence of iron in the N lobe. Our results indicate that *Neisseria* cannot use the entire serum TF iron supply and that the primary function of TbpB is to select and concentrate on the cell surface only those forms of TF that are able to provide iron to the bacterium.

Because TbpB primarily binds the C lobe of hTF through its N lobe, we performed steered molecular docking for the TbpB–hTF complex based on previous docking studies for the porcine complex²⁰ and on our mutagenesis results. We collected small-angle X-ray scattering (SAXS) data on the TbpB–(holo)hTF complex (Supplementary Figs 12 and 13) and used GASBOR²⁴ to construct the SAXS envelope (Fig. 3c and Supplementary Fig. 13). The resulting molecular envelope describes the spatial arrangement of TbpB and hTF, and was used to fit the TbpB–hTF complex structure. Binding TbpB to hTF buries ~1,300 Å² of surface area and, notably, uses a region of the hTF C lobe distinct from the site where TbpA binds.

Structure of the triple complex by single-particle EM

On the basis of the TbpA–(apo)hTF crystal structure and the SAXS solution structure of the TbpB–(holo)hTF complex, we formed an *in silico* model for the TbpA–TbpB–(holo)hTF triple complex by

superposing the two complexes along the C1 subdomain of hTF (Fig. 4a). To test this model, we assembled the triple complex from its components (Fig. 4b) and visualized the resulting particles by negative staining electron microscopy (EM) (Fig. 4c and Supplementary Fig. 14). A set of 4,240 particles was subjected to a reference-free classification to identify subsets of like images, representing molecules viewed in the same orientation; the members of each class were then averaged to suppress noise. Several of the class averages show a central density, ~45 Å across, to which two small globular densities are appended at points about 120° apart around its periphery (for example, Fig. 4c (asterisk) and Supplementary Fig. 14). A plausible interpretation is that the central density corresponds to the β-barrel domain of TbpA and the two appended densities to TbpB and hTF (Supplementary Fig. 14), in agreement with our model for the triple complex.

Iron extraction and import

The X-ray, SAXS and EM structures support a consistent arrangement for the TbpA–TbpB–(holo)hTF complex. Although TbpA and TbpB each bind hTF tightly through the C lobe, they have unique, non-overlapping binding sites (Fig. 5a). A consequence of the assembly of the triple complex is the formation of an enclosed chamber (volume of ~1,000 Å³) at the union of the three protein components, which sits directly above the plug domain of TbpA (Fig. 5b). This chamber may serve two important roles for iron acquisition by the bacteria: (1) prevent diffusion of iron released from hTF; and (2) guide the iron towards the β-barrel domain of TbpA for subsequent transport.

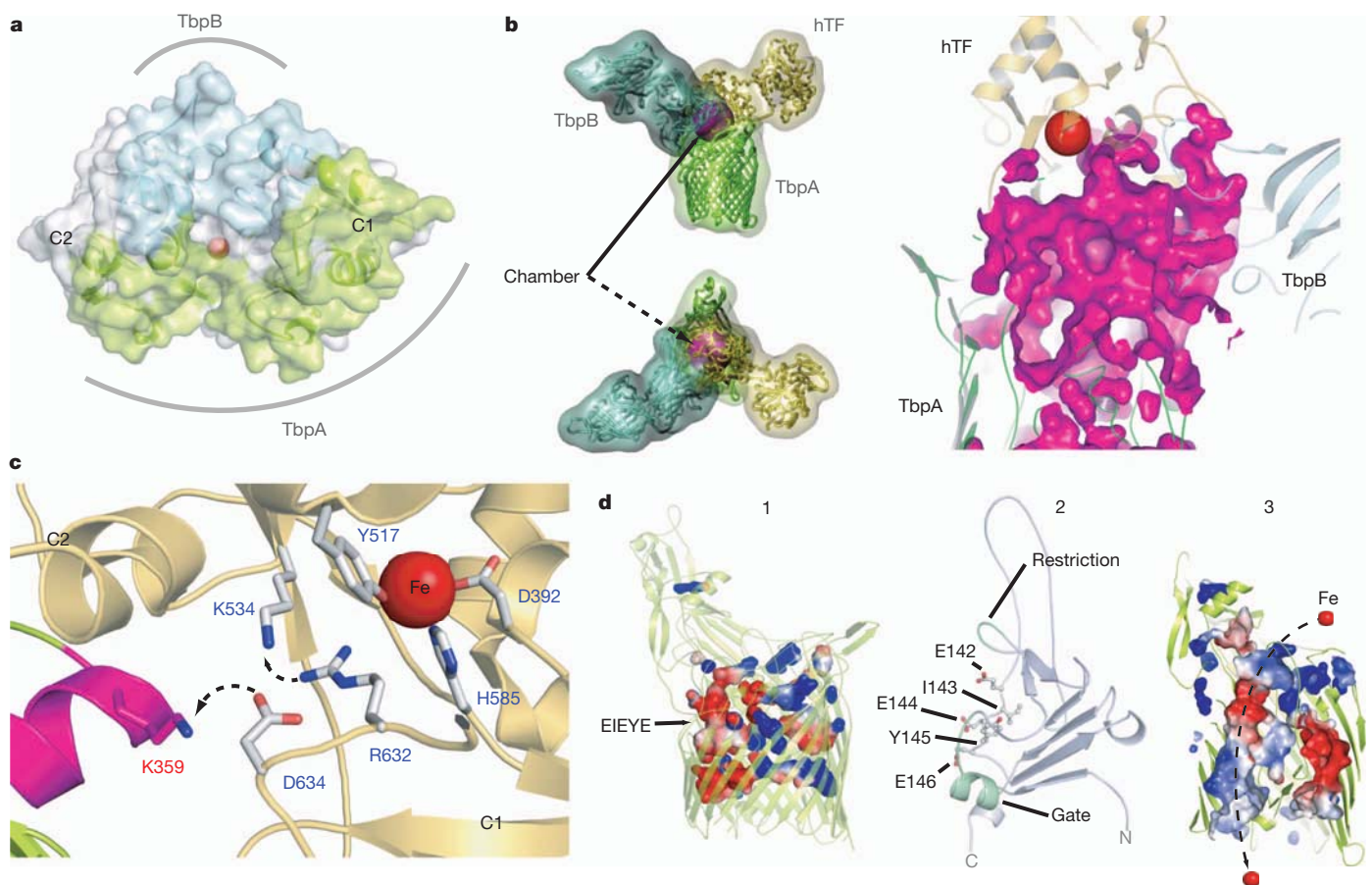


Figure 5 | Mechanism for iron import. **a**, Binding surfaces of TbpA (green) and TbpB (cyan) mapped onto the hTF C lobe. **b**, Enclosed chamber formed by TbpA–TbpB–(holo)hTF (left, magenta sphere). A cutaway view (right) from inside the chamber illustrates the proximity of the iron (red). **c**, Model for iron release. Conserved K359 in the L3 helix finger is positioned to interact with

residues that regulate iron release in eukaryotic iron uptake. **d**, Import of iron through TbpA. 1, an electrostatic surface depicts cavities between the TbpA barrel and plug domain; 2, plug domain constrictions close the tunnel; 3, molecular dynamics simulations show removal of constrictions upon interaction with TonB.

A plausible mechanism for iron extraction from hTF is shown in Fig. 5c. Insertion of the TbpA L3 helix finger into the cleft between the C1 and C2 subdomains of hTF positions a conserved lysine (TbpA K359; Supplementary Fig. 4) near the hTF triad of charged residues (hTF K534, R632 and D634) that has been implicated in iron release from the C lobe²⁵. TbpA K359 is perfectly situated to interact with D634, which would disrupt the charge neutralization it normally provides to the two basic triad residues K534 and R632. This potential charge repulsion between the hTF C1 and C2 subdomains could induce cleft opening (as observed in the TbpA-(apo)hTF crystal structures), resulting in distortion of the C-lobe iron-binding site and subsequent iron release. Notably, a recent study indicates that an hTF D634A mutant has a rate of iron release that is 80-fold faster at pH 5.6 than the control under the same conditions²⁶.

To investigate iron transport across the outer membrane, steered molecular dynamics was used to simulate interactions between TbpA and TonB (Supplementary Fig. 15). In the ground state structure, a large, highly negative transmembrane cavity is located between the barrel wall and the plug domain, but access is restricted on the extracellular side by residues 91–96 (restriction loop) and on the periplasmic side by residues 65–71 (helical gate) from the plug domain (Fig. 5d). When force (designed to mimic interaction with TonB) is applied to the plug domain, it sequentially unfolds beginning with removal of the helical gate followed by the restriction loop, producing an unobstructed pathway from the extracellular space to the periplasm (Supplementary Movie 1). This pathway is lined by the EIEYE motif of the plug domain¹³, which contains multiple oxygen donor groups that could transiently bind iron as it is transported through TbpA.

Concluding remarks

Humans and bacteria have each developed unique strategies to acquire iron from serum transferrin^{27,28} (Supplementary Fig. 16). Our TbpA–TbpB–hTF X-ray, SAXS and EM structures indicate a mechanism for bacterial uptake of iron with the following characteristics: (1) a large TbpA–hTF binding interface with many human-TF-specific interactions; (2) iron removal from the hTF C lobe by insertion of a helical element from TbpA into the iron-binding cleft; and (3) iron transport across the outer membrane after TonB-dependent conformational changes in the TbpA plug domain. This system allows efficient extraction of iron despite the extremely high affinity of hTF bound iron at neutral pH. The TbpB co-receptor, which is tethered to the cell surface by a long, unstructured polypeptide chain, is able to attract and preferentially bind hTF with iron in the C lobe, thereby increasing the efficiency of the system. Crucial to the mechanism, TbpA and TbpB associate with different regions of the hTF C lobe, creating an enclosed chamber above the plug domain to ensure that iron is efficiently sequestered and directionally transported through the TbpA barrel (Supplementary Movie 2). Finally, as TbpA and TbpB are surface-exposed, antigenic and required for neisserial infections²⁹, our structures provide the necessary information for structure-based vaccine and drug design³⁰.

METHODS SUMMARY

The TbpA–(apo)hTF complex was crystallized from TbpA expressed in *Escherichia coli* and apo-hTF purchased from Sigma-Aldrich. For the TbpA–(apo)hTF C-lobe structure, hTF incorporating a protease cleavage site between N and C lobes was expressed in BHK cells and purified as described³¹. Full-length N-His-tagged hTFs including holo, authentic apo, and both monoferric forms were expressed in BHK cells and purified as described²³. The diferric hTF structure was solved using protein purchased from Sigma-Aldrich. The hTF C-lobe structure was solved using hTF C lobe from a construct containing a TEV protease site between the N and C lobes, expressed in BHK cells, and purified³². TbpB was expressed in *E. coli*. X-ray data were collected at GM/CA and SER-CAT beamlines of the Advanced Photon Source synchrotron. SAXS data was collected on beamline BL4-2 at the Stanford Synchrotron Radiation Lightsource. EM data were collected on a CM120–LaB6 electron microscope (FEI), operating at 120 kV. Molecular dynamics simulations were performed using the program NAMD³³.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 19 September 2011; accepted 9 January 2012.

Published online 12 February 2012.

- Deasy, A. & Read, R. C. Challenges for development of meningococcal vaccines in infants and children. *Expert Rev. Vaccines* **10**, 335–343 (2011).
- Centers for Disease Control and Prevention. Cephalosporin susceptibility among *Neisseria gonorrhoeae* isolates—United States, 2000–2010. *Morbid. Mortal. Weekly Rep.* **60**, 873–877 (2011).
- Grifantini, R. *et al.* Identification of iron-activated and -repressed Fur-dependent genes by transcriptome analysis of *Neisseria meningitidis* group B. *Proc. Natl Acad. Sci. USA* **100**, 9542–9547 (2003).
- Noiraj, N., Guillier, M., Barnard, T. J. & Buchanan, S. K. TonB-dependent transporters: regulation, structure, and function. *Annu. Rev. Microbiol.* **64**, 43–60 (2010).
- Boulton, I. C. *et al.* Transferrin-binding protein B isolated from *Neisseria meningitidis* discriminates between apo and diferric human transferrin. *Biochem. J.* **334**, 269–273 (1998).
- Krell, T. *et al.* Insight into the structure and function of the transferrin receptor from *Neisseria meningitidis* using microcalorimetric techniques. *J. Biol. Chem.* **278**, 14712–14722 (2003).
- Anderson, J. E., Sparling, P. F. & Cornelissen, C. N. Gonococcal transferrin-binding protein 2 facilitates but is not essential for transferrin utilization. *J. Bacteriol.* **176**, 3162–3170 (1994).
- Irwin, S. W., Averil, N., Cheng, C. Y. & Schryvers, A. B. Preparation and analysis of isogenic mutants in the transferrin receptor protein genes, *tbpA* and *tbpB*, from *Neisseria meningitidis*. *Mol. Microbiol.* **8**, 1125–1133 (1993).
- Rokbi, B. *et al.* Evaluation of recombinant transferrin-binding protein B variants from *Neisseria meningitidis* for their ability to induce cross-reactive and bactericidal antibodies against a genetically diverse collection of serogroup B strains. *Infect. Immun.* **65**, 55–63 (1997).
- Weynants, V. E. *et al.* Additive and synergistic bactericidal activity of antibodies directed against minor outer membrane proteins of *Neisseria meningitidis*. *Infect. Immun.* **75**, 5434–5442 (2007).
- Price, G. A., Masri, H. P., Hollander, A. M., Russell, M. W. & Cornelissen, C. N. Gonococcal transferrin binding protein chimeras induce bactericidal and growth inhibitory antibodies in mice. *Vaccine* **25**, 7247–7260 (2007).
- Yost-Daljev, M. K. & Cornelissen, C. N. Determination of surface-exposed, functional domains of gonococcal transferrin-binding protein A. *Infect. Immun.* **72**, 1775–1785 (2004).
- Noto, J. M. & Cornelissen, C. N. Identification of TbpA residues required for transferrin-iron utilization by *Neisseria gonorrhoeae*. *Infect. Immun.* **76**, 1960–1969 (2008).
- Wally, J. *et al.* The crystal structure of iron-free human serum transferrin provides insight into inter-lobe communication and receptor binding. *J. Biol. Chem.* **281**, 24934–24944 (2006).
- Cornelissen, C. N., Biswas, G. D. & Sparling, P. F. Expression of gonococcal transferrin-binding protein 1 causes *Escherichia coli* to bind human transferrin. *J. Bacteriol.* **175**, 2448–2450 (1993).
- Schryvers, A. B. & Morris, L. J. Identification and characterization of the transferrin receptor from *Neisseria meningitidis*. *Mol. Microbiol.* **2**, 281–288 (1988).
- Stokes, R. H., Oakhill, J. S., Joannou, C. L., Gorrings, A. R. & Evans, R. W. Meningococcal transferrin-binding proteins A and B show cooperation in their binding kinetics for human transferrin. *Infect. Immun.* **73**, 944–952 (2005).
- Schryvers, A. B. & Gonzalez, G. C. Comparison of the abilities of different protein sources of iron to enhance *Neisseria meningitidis* infection in mice. *Infect. Immun.* **57**, 2425–2429 (1989).
- Calmettes, C. *et al.* Structural variations within the transferrin binding site on transferrin-binding protein B, TbpB. *J. Biol. Chem.* **286**, 12683–12692 (2011).
- Moraes, T. F., Yu, R. H., Strynadka, N. C. & Schryvers, A. B. Insights into the bacterial transferrin receptor: the structure of transferrin-binding protein B from *Actinobacillus pleuropneumoniae*. *Mol. Cell* **35**, 523–533 (2009).
- Cornelissen, C. N., Anderson, J. E. & Sparling, P. F. Characterization of the diversity and the transferrin-binding domain of gonococcal transferrin-binding protein 2. *Infect. Immun.* **65**, 822–828 (1997).
- Silva, L. P. *et al.* Conserved interaction between transferrin and transferrin-binding proteins from porcine pathogens. *J. Biol. Chem.* **286**, 21353–21360 (2011).
- Mason, A. B. *et al.* Expression, purification, and characterization of authentic monoferric and apo-human serum transferrins. *Protein Expr. Purif.* **36**, 318–326 (2004).
- Svergun, D. I., Petoukhov, M. V. & Koch, M. H. Determination of domain structure of proteins from X-ray solution scattering. *Biophys. J.* **80**, 2946–2953 (2001).
- Halbrooks, P. J. *et al.* Investigation of the mechanism of iron release from the C-lobe of human serum transferrin: mutational analysis of the role of a pH sensitive triad. *Biochemistry* **42**, 3701–3707 (2003).
- Steere, A. N., Byrne, S. L., Chasteen, N. D. & Mason, A. B. Kinetics of iron release from transferrin bound to the transferrin receptor at endosomal pH. *Biochim. Biophys. Acta* <http://dx.doi.org/10.1016/j.bbagen.2011.06.003> (2011).
- Cheng, Y., Zak, O., Aisen, P., Harrison, S. C. & Walz, T. Structure of the human transferrin receptor-transferrin complex. *Cell* **116**, 565–576 (2004).
- Eckenroth, B. E., Steere, A. N., Chasteen, N. D., Everse, S. J. & Mason, A. B. How the binding of human transferrin primes the transferrin receptor potentiating iron release at endosomal pH. *Proc. Natl Acad. Sci. USA* **108**, 13089–13094 (2011).

29. Hobbs, M. M. *et al.* Experimental gonococcal infection in male volunteers: cumulative experience with *Neisseria gonorrhoeae* strains FA1090 and MS11mkC. *Front. Microbiol.* **2**, 123 (2011).
30. Scarselli, M. *et al.* Rational design of a meningococcal antigen inducing broad protective immunity. *Sci. Transl. Med.* **3**, 91ra62 (2011).
31. Zak, O. & Aisen, P. A new method for obtaining human transferrin C-lobe in the native conformation: preparation and properties. *Biochemistry* **41**, 1647–1653 (2002).
32. Steere, A. N. *et al.* Properties of a homogeneous C-lobe prepared by introduction of a TEV cleavage site between the lobes of human transferrin. *Protein Expr. Purif.* **72**, 32–41 (2010).
33. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements N.N., N.C.E., M.O., E.B. and S.K.B. are supported by the Intramural Research Program of the NIH, National Institute of Diabetes and Digestive and Kidney Diseases. M.O. was initially funded by an EPSRC Research Committee Studentship awarded to S.K.B. and R.W.E. N.M. and A.C.S. are supported by the Intramural Research Program of the NIH, National Institute of Arthritis and Musculoskeletal and Skin Diseases. A.B.M. was supported in part by USPHS grant R01-DK21739. A.N.S. is funded by an AHA Predoctoral Fellowship (10PRE4200010). E.T. acknowledges NIH support by R01-GM086749, U54-GM087519 and P41-RR05969. All the simulations were performed using TeraGrid resources (MCA06N060). We thank the respective staffs at the Southeast Regional Collaborative Access Team (SER-CAT) and General Medicine and Cancer Institutes Collaborative Access Team (GM/CA-CAT) beamlines at the Advanced Photon Source, Argonne National Laboratory for their assistance during data

collection. Use of the Advanced Photon Source was supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. W-31-109-Eng-38 (SER-CAT), and by the US Department of Energy, Basic Energy Sciences, Office of Science, under contract No. DE-AC02-06CH11357 (GM/CA-CAT). Portions of this research were carried out at the Stanford Synchrotron Radiation Laboratory, a national user facility operated by Stanford University on behalf of the US Department of Energy, Office of Basic Energy Sciences. The SSRL Structural Molecular Biology Program is supported by the Department of Energy, Office of Biological and Environmental Research, and by the National Institutes of Health, National Center for Research Resources, Biomedical Technology Program.

Author Contributions N.N., N.C.E., M.O. and S.K.B. expressed, purified and crystallized TbpA, TbpB and various hTFs. N.N. solved all crystal structures and the SAXS structure and analysed all data. A.B.M. and A.N.S. designed and purified apo-hTF, holo-hTF, hTF-Fe_N and hTF-Fe_C for binding experiments with TbpA and TbpB; they also expressed and purified hTF C lobe for the corresponding structure (PDB code 3SKP). P.A. and O.Z. expressed and purified hTF C lobe for the TbpA-(apo)hTF C-lobe structure (PDB code 3V89). N.M. and A.C.S. designed, conducted and analysed EM experiments. E.T. and J.G. designed, conducted and analysed molecular dynamics simulations. E.B. participated in the data collection and analysis of the SAXS data. R.W.E., A.R.G. and S.K.B. conceived and designed the original project. N.N. and S.K.B. wrote the manuscript.

Author Information Coordinates and structure factors for TbpA-(apo)hTF, TbpA-(apo)hTF C lobe, diferric hTF, apo-hTF C lobe and TbpB are deposited in the Protein Data Bank under accession codes 3V8X, 3V89, 3V83, 3SKP and 3V8U, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.K.B. (skbuchan@helix.nih.gov).

METHODS

Cloning, expression and purification of TbpA. The *N. meningitidis* TbpA sequence from strain K454 (B15:P1.7,16) was subcloned into pET20b (Novagen) containing an N-terminal 10×-His tag. TbpA mutants were created using site-directed mutagenesis using QuikChange (Stratagene). For structural studies, mutation of M889 to Tyr improved expression levels. TbpA was expressed in BL21(DE3) cells at 20 °C without induction in terrific broth (TB) and carbenicillin. Expression for the mutants followed the same protocol.

For purification, cells were re-suspended in lysis buffer (50 mM Tris-HCl, pH 7.5, 200 mM NaCl, 1 mM MgCl₂, 10 µg ml⁻¹ DNaseI, 100 µg ml⁻¹ AEBSF) and lysed by two passages through an Emulsiflex C3 (Avestin) homogenizer at 4 °C. The lysate was centrifuged at 12,000g for 10 min to remove unlysed cells and the supernatant was incubated with 2% Triton X-100 for 30 min at room temperature. The mixture was centrifuged at 160,000g for 90 min at 4 °C. The membrane pellets were re-suspended in 50 mM Tris-HCl, pH 7.5, 200 mM NaCl, 20 mM imidazole and solubilized by constant stirring using 5% Elugent for 16 h at 4 °C. Solubilized membranes were centrifuged at 265,000g for 60 min at 4 °C and the supernatant filtered and applied to a 15-ml Ni-NTA column. TbpA was eluted using 250 mM imidazole. Peak fractions were concentrated and applied to an S-300HR Sephacryl size exclusion column (GE Healthcare) using 20 mM Tris-HCl, pH 7.5, 200 mM NaCl, 0.8% C₈E₄ and 0.02% NaN₃. Peak fractions were verified using SDS-PAGE and western blot analysis using an anti-His monoclonal antibody (Sigma).

Cloning, expression and purification of TbpB. The TbpB sequence (starting at residue L22) from *N. meningitidis* K454 was codon optimized and synthesized by GenScript and subsequently subcloned into a pET28b vector (Novagen). TbpB was expressed in T7-Express cells (NEB) at 37 °C with IPTG induction at an optical density at 600 nm of 0.75–1.0 with continued expression for 4 h. Mutants were expressed using the same protocol.

For purification, cells were harvested and re-suspended in 5 ml PBS per gram of cell paste and supplemented with 10 µg ml⁻¹ AEBSF and 100 µg ml⁻¹ DNaseI. Cells were lysed by French press and then centrifuged for 45 min at 38,400g. The supernatant was applied to a Ni-NTA column and washed with 10 column volumes of PBS. A final wash was performed with PBS containing 20 mM imidazole before elution with PBS/250 mM imidazole. Eluted protein was then dialysed against PBS overnight at 4 °C. For constructs where the His tag was removed, TEV-HIS protease was added, the sample was dialysed and then passed through a second Ni-NTA column to remove uncleaved protein and protease. Finally, samples were purified by size-exclusion chromatography in PBS/0.02% NaN₃.

Crystallization and data collection. For crystallization of the TbpA-hTF complex, apo-human transferrin (Sigma) was mixed with TbpA at a 2:1 ratio and incubated on ice for 1 h. The complex was isolated using Sephacryl S300HR chromatography equilibrated with 20 mM Tris-HCl, pH 7.5, 200 mM NaCl, 10 mM Na-citrate, 1 mM EDTA, 0.8% C₈E₄ (Anatrace) and 0.02% NaN₃. Fractions corresponding to the TbpA-hTF complex were verified using SDS-PAGE, pooled and concentrated to 10 mg ml⁻¹. Heptane-1,2,3-triol was added to 3% final concentration, incubated on ice for 30 min and then the protein sample was filtered before crystallization. Sparse matrix screening was performed using a TTP Labtech Mosquito crystallization robot using hanging-drop vapour diffusion and plates incubated at 21 °C. The best crystals were grown in 24-well Linbro plates (Hampton Research) from 20% Peg3350 and 200 mM BaBr₂. Data were collected at the SER-CAT beamline of the Advanced Photon Source of Argonne National Laboratory and data processed using HKL2000³⁴. The space group was *P*₂₁₂₁₂₁ with one mol per asymmetric unit (ASU) and final cell parameters *a* = 91.014, *b* = 129.362, *c* = 198.589, α = 90.00, β = 90.00, γ = 90.00.

For crystallization of the TbpA-hTF C-lobe complex, hTF C lobe³¹ was mixed in a 2:1 ratio with TbpA and the complex isolated by Sephacryl S300HR chromatography using 20 mM Tris-HCl, pH 7.5, 200 mM NaCl, 0.1% LDAO and 0.02% NaN₃. Final crystal conditions consisted of 21% PEG 1000, 100 mM sodium acetate buffer (pH 4.8), 200 mM NaCl, 0.1% LDAO and 3% heptane-1,2,3-triol. Data were collected and processed as described for the TbpA-hTF complex. The space group was *P*₂₁ with one mol per ASU with final cell parameters *a* = 58.055, *b* = 107.592, *c* = 130.721, α = 90.00, β = 94.48, γ = 90.00.

TbpB was crystallized from a 10 mg ml⁻¹ solution with 2.0 M NaCl and 2.0 M ammonium sulphate. Data were collected at the GM/CA CAT beamline of the Advanced Photon Source of Argonne National Laboratory and data were processed using HKL2000³⁴. The space group was *P*₂₁ with two molecules per ASU with final cell parameters *a* = 75.288, *b* = 82.761, *c* = 111.882, α = 90.00, β = 105.95, γ = 90.00.

For diferric hTF crystallization, 100 mg of holo-hTF (Sigma) was solubilized and further purified by Sephacryl S300HR chromatography using 20 mM Tris-HCl, pH 7.5 and 200 mM NaCl. The protein was then concentrated to ~50 mg ml⁻¹

and crystallized using 100 mM HEPES, pH 7.5, 1.6 M ammonium sulphate, and 2% PEG 1000, with red-tinted crystals appearing only after several months and being extremely sensitive to even slight temperature changes. Drops containing the crystals were quickly hydrated with 3.4 M ammonium sulphate immediately before being flash cooled in liquid nitrogen and stored for data collection. Data were collected at the SER-CAT beamline of the Advanced Photon Source of Argonne National Laboratory and data were processed using HKL2000³⁴. The space group is *C*₂ with six molecules per ASU and final cell parameters *a* = 254.53, *b* = 173.00, *c* = 150.15, α = 90.00, β = 123.26, γ = 90.00.

For hTF C-lobe crystallization, (holo)C lobe³² was mixed with excess TbpB N lobe and the complex isolated by size-exclusion chromatography as above in 25 mM Tris pH 8.0, 200 mM NaCl. The complex was concentrated to ~10 mg ml⁻¹ and broad screening performed using a Mosquito crystallization robot. Several crystallization conditions were observed; however, none was red in colour as might be expected for iron-bound crystals and most contained citrate, which is a known iron chelator. Data were collected and analysed as for TbpB. The space group was *I*₄₂₂ with 1 mol of hTF C lobe per ASU with final cell parameters *a* = 95.847, *b* = 95.847, *c* = 204.140, α = 90.00, β = 90.00, γ = 90.00. No TbpB N lobe was present in the crystals.

Structure determination. For TbpA-hTF, we were unable to collect useful heavy atom derivatives for experimental phasing, and selenomethionine-substituted TbpA protein yields were not sufficient for crystallization. We eventually used molecular replacement in PHASER-CCP4³⁵ to solve the TbpA-hTF complex structure. Here, we first searched for each of the two domains (N lobe and C lobe) of hTF using the deposited coordinates (PDB code 2HAV), which produced good solutions with *Z*-scores above 8. However, although the electron density for the hTF molecule was reasonable, the electron density for TbpA was poor and could not be used for model building. Our attempts at using known TonB-dependent transporter structures as search models (barrel and plug, together and individually) were unsuccessful (low *Z*-scores and LLG scores). We then aligned the TbpA sequence to our structure-based sequence alignment reported in our recent review⁴ and found that TbpA contained many conserved regions characteristic of TonB-dependent transporters. Using the alignment between TbpA and its closest relative, FhuA (10% identity, ClustalW), and trimming the extracellular loops, 500 models within a root mean squared deviation of 5 Å were produced using Modeller (Accelrys). Each of these models was then used for molecular replacement within PHASER-CCP4³⁵, with two of them producing *Z*-scores above 8. The solution with the highest LLG (containing both hTF and the TbpA model) was refined in PHENIX³⁶ producing an initial *R*/*R*_{free} value of 0.43/0.48. Further model building was performed using COOT³⁷ and subsequent refinement done in PHENIX³⁶ and BUSTER-TNT³⁸. During the final states of refinement, extra density was observed which was mapped to residues N413 and N611 of hTF, both of which are reported as possible N-linked glycosylation sites. Therefore, N-linked glycans were built for these two residues. The final structure was solved to 2.60 Å with *R*/*R*_{free} values of 0.22/0.28. The TbpA-hTF C-lobe structure was solved by molecular replacement using the coordinates from the TbpA-hTF (full-length) structure reported here. Two search models were formed, one for TbpA and one for hTF C lobe. PHASER-CCP4³⁵ was used for molecular replacement and subsequent refinement performed using PHENIX³⁶. The structure was solved to 3.1 Å resolution with final *R*/*R*_{free} values of 0.24/0.29.

The TbpB structure was solved by molecular replacement using PDB code 3HOL. An initial model was created using the Swiss Model server³⁹ that was subsequently divided into four different search domains. PHASER-CCP4³⁵ was used for molecular replacement and subsequent refinement performed using PHENIX³⁶. The structure was solved to 2.40 Å resolution with final *R*/*R*_{free} values of 0.25/0.30.

The diferric hTF crystal structure was solved by molecular replacement using Phaser-CCP4³⁵. Search models for the N lobe and C lobe were created separately with the program Chainsaw (CCP4) using the existing diferric porcine TF coordinates (PDB code 1H76). Six copies of each lobe (six molecules of hTF total) were found in the ASU and the iron sites were easily observed in the difference density. These iron sites were further verified in an anomalous difference electron density map. Refinement was performed using PHENIX³⁶ and the structure was solved to 2.1 Å resolution with final *R*/*R*_{free} values of 0.19/0.23.

The non-glycosylated hTF-C-lobe structure was solved by molecular replacement using PDB code 2HAU. An initial search model was formed by truncating the N-lobe domain. PHASER-CCP4³⁵ was used for molecular replacement and subsequent refinement performed using PHENIX³⁶. The structure was solved to 1.7 Å resolution with final *R*/*R*_{free} values of 0.17/0.19. For all structures, figures were made with PyMOL (Schrodinger) or Chimera⁴⁰ and annotated and finalized with Adobe Illustrator.

Dot blots. Whole cells ($2\ \mu\text{l}$, $0.01\ \text{g ml}^{-1}$) and cell lysates (unmodified for TbpB samples, or incubated for 3 h with 2 mM EDTA and 1% DDM at room temperature for TbpA samples) were spotted onto nitrocellulose membrane and allowed to dry at room temperature. The membranes were then blocked with PBST 2% BSA for 15 min, washed and probed with HRP-conjugated hTF (1:1,000) (Jackson ImmunoResearch) for 15 min. The membrane was then washed and imaged using the colorimetric substrate 3,3'-diaminobenzidine (Sigma) where the appearance of a red dot indicated specific binding of the hTF–HRP conjugate. The results from the mutants were compared to wild-type Tbp to determine their effect on hTF binding.

ELISA. Whole cells ($100\ \mu\text{l}$ at $10\ \text{mg ml}^{-1}$ or $1\ \text{mg ml}^{-1}$ in PBS) of wild-type TbpA, empty vector control (pET20b) and TbpA mutants were added to a NUNC polystyrene 96-well plate (Fisher Scientific) and incubated at 37°C overnight. Wells were washed $2\times$ with PBST and then blocked with PBST 2% BSA for 30 min and probed with hTF–HRP (1:1,000) for 15 min. Wells were washed $2\times$ in PBST, $2\times$ in PBS, and then developed using $100\ \mu\text{l}$ 3,3',5,5'-tetramethylbenzidine substrate (TMB, Sigma) for 5 min and terminated using Stop solution (Sigma). Absorbances of each well were determined using a BioRad iMark plate reader at 450 nm and data normalized and compared to wild-type TbpA. Each experiment was performed in triplicate and data reported with standard errors.

Antibody blocking assays. Using the TbpA–hTF crystal structure reported here, we designed four different peptides based on four loops from TbpA (loops 3, 7, 11 and plug loop) to be used as antigens for polyclonal antibody development (Precision Antibody). A fifth polyclonal antibody was developed using purified full-length TbpA ($1\times$ PBS 7.4, 0.1% DDM). An ELISA was designed to probe whether or not these antibodies could block hTF binding. Here, TbpA–His (20 ng) was incubated for 20 min in a final volume of $100\ \mu\text{l}$ either alone or in the presence of each antibody (1:20) individually in PBS containing 0.05% Cymal-6 (Anatrace). In addition, we tested the antibodies that targeted TbpA loops in combinations to determine if an additive effect could be observed. Each sample was then transferred to a 96-well Ni-NTA Agarose HiSorb plate (Qiagen) and incubated for 30 min and washed $2\times$ with PBST + 0.05% Cymal-6. Assays were performed as described in the previous section. In a second set of ELISAs, TbpA–His was first bound to the Ni-NTA Agarose HiSorb plate before incubation with antibodies. Results were analysed and initial graphs made using Microsoft Excel. The graphs were then imported, annotated and finalized with Adobe Illustrator.

Protease accessibility of TbpA and TbpA mutants. To confirm that TbpA and the TbpA mutants were being properly presented at the surface of the bacteria, we treated whole cells with trypsin ($5\ \mu\text{g ml}^{-1}$ final concentration) for 15 min at room temperature and the reaction was stopped by the addition of AEBSF ($0.2\ \text{mg ml}^{-1}$ final concentration). The cells were then centrifuged and supernatant removed. The pellets were then re-suspended in LDS loading buffer, boiled for 10 min, centrifuged for 10 min, and then separated on a NuPAGE Novex 4–12% Bis-Tris gel. The samples were then transferred to a PVDF membrane using an iBlot system (Invitrogen) and western blot analysis performed using a polyclonal anti-TbpA antibody (1:1,000) and a monoclonal anti-His HRP-conjugated antibody (Sigma) (1:5,000). Here, each membrane was blocked with 2% BSA in $1\times$ PBST for 15 min and then probed with either the anti-TbpA or anti-His HRP conjugated antibody for 30 min. The anti-TbpA membrane was then washed $2\times$ with 2% BSA in $1\times$ PBST and then probed with anti-mouse HRP-conjugated secondary antibodies for an additional 30 min. Both membranes were then washed $2\times$ with $1\times$ PBST, $2\times$ with $1\times$ PBS, and then imaged using the colorimetric substrate 3,3'-diaminobenzidine (Sigma). The results from the mutants were compared to wild-type TbpA to determine which constructs were being presented on the surface of the cells.

Transferrin competition assays. To determine if the affinity of hTF to either TbpA or TbpB is affected by the conformation or coordination state of the N lobe, we performed an ELISA competition assay using apo-hTF, holo-hTF, hTF–Fe_N (iron bound in N lobe only) and hTF–Fe_C (iron bound in C lobe only), which were expressed in BHK cells and purified as described²³. Here, His–TbpA (20 ng) or His–TbpB (20 ng) was incubated for 15 min in a final volume of $100\ \mu\text{l}$ in a 96-well Ni-NTA HiSorb plate (Qiagen) in $1\times$ PBS (0.05% cymal-6 was added to all buffers for His–TbpA). Wells were washed $2\times$ with PBST, blocked with 2% BSA in PBST for 15 min, and then incubated with each of the transferrin constructs (apo, holo, Fe_N, Fe_C) (100 ng each) for 15 min. Wells were washed again $2\times$ with PBST and then probed for 20 min with HRP-conjugated hTF (1:1,000) (Jackson ImmunoResearch) in $100\ \mu\text{l}$ final volume. Wells were washed again $2\times$ with PBST and then $2\times$ with PBS and imaged using $100\ \mu\text{l}$ 3,3',5,5'-tetramethylbenzidine substrate (Sigma) for ~ 5 min and terminated using Stop solution (Sigma). Data were collected and analysed as described above.

Small-angle X-ray scattering analysis. The TbpB–hTF complex was dialysed overnight at 4°C into TBS, pH 7.4 (25 mM Tris, 137 mM NaCl, 3 mM KCl) and

then filtered using a $0.2\ \mu\text{m}$ spin filter. Data were collected at concentrations of 1, 2.5 and $5\ \text{mg ml}^{-1}$ at Stanford Synchrotron Radiation Lightsource beamline BL4-2. Data reduction and analysis were performed using the beamline software SASAtool. The program AutoGNOM of the ATSAS suite⁴¹ was used to generate $P(r)$ curves and to determine maximum dimension (D_{max}) and radius of gyration (R_g) from the scattering intensity curve ($I(q)$ versus q) in an automatic, unbiased manner, and rounds of manual fitting in GNOM⁴² were used to verify these values. *Ab initio* molecular envelopes were computed by the programs GASBOR²⁴. Ten iterations of GASBOR were averaged using DAMAVER⁴³. Docking of the TbpB and diferric hTF crystal structures into the molecular envelope was performed manually, guided by both previous docking studies²⁰ and mutagenesis results. Figures were made with PyMOL and annotated and finalized with Adobe Illustrator.

Modelling the TbpA–TbpB–hTF triple complex. The *in silico* TbpA–TbpB–hTF triple complex was assembled based on our crystal structures (TbpA–(apo)hTF, diferric hTF, TbpB) and SAXS analysis (TbpB–(holo)hTF) reported here. The crystal structure (TbpA–hTF) was aligned with our TbpB–hTF model using the C1 subdomain of hTF as a reference, yielding a triple complex containing a 1:1:1 ratio of TbpA, TbpB and hTF. Figures were made with PyMOL and/or Maya (Autodesk) and annotated and finalized with Adobe Illustrator.

Electron microscopic analysis. The triple complex (TbpA–TbpB–(holo)hTF) was prepared from separately purified components by first forming a complex between TbpB and (holo)hTF, which was purified by size-exclusion chromatography in $1\times$ PBS. Cymal-6 was added to a final concentration of 0.05% and purified TbpA (in $1\times$ PBS, 0.05% Cymal 6) was added to the mixture using an excess of the TbpB–(holo)hTF complex. The triple complex (which retains a 1:1:1 stoichiometry) was isolated by sized exclusion chromatography in buffer A ($1\times$ PBS in 0.05% Cymal 6) and used immediately for EM experiments. The complex was diluted with buffer A to an optimal concentration for EM (determined empirically to be $\sim 1\ \mu\text{g ml}^{-1}$). Drops ($4\ \mu\text{l}$ each) were applied to carbon-coated, glow-discharged EM grids (EMS). After 1 min, the grid was blotted, washed twice with buffer A, once with distilled water, and then stained with 2% uranyl acetate. Grids were observed with a CM120 La-B6 electron microscope (FEI), operating at 120 kV. Micrographs were recorded on SO163 film (Kodak) at a nominal magnification of 45,000, and digitized on a Nikon Coolscan 9000 at a rate corresponding to $1.55\ \text{\AA}$ per pixel. The large majority of particles distributed evenly on the grid and were essentially uniform in size (~ 90 – $110\ \text{\AA}$ in diameter), indicative of a homogeneous population.

The particles were variable in substructure, suggesting that the molecules deposit on the grid in a variety of orientations. Accordingly, a data set of 4,240 particles was subjected to a 'reference free' classification, using SPIDER⁴⁴, EMAN⁴⁵ and Bsoft⁴⁶. Images were picked using a 256×256 pixel box, and binned four times (to $6.2\ \text{\AA}$ per pixel) to increase the signal-to-noise ratio and the speed of calculation. Initial reference-free classification and averaging were performed using EMAN; further classification was done in SPIDER, using principal component analysis (PCA) with three cycles of iteration. We chose to obtain 56 final class averages, based on a cluster distribution obtained from PCA.

The coordinates of the modelled triple complex were converted to a density map and low-pass filtered to $15\ \text{\AA}$. The sampling rate of the density map was set to be same as the EM images and two-dimensional projections were calculated at angular increments of 30° (Supplementary Fig. 14b). Comparisons and matching between the EM class averages and the model re-projections were done in terms of cross-correlation coefficients. A few ambiguous cases were reassigned by visual assessment. Figures were made with Chimera and annotated and finalized with Adobe Illustrator.

Molecular dynamics simulations. For simulations of TbpA bound to apo-hTF, a membrane-water system containing the protein complex was first built using VMD⁴⁷. The complex was placed in a DMPE bilayer as used previously⁴⁸, with the barrel of TbpA aligned with the membrane's hydrophobic core, and then fully solvated. Disulphide bonds for three pairs in TbpA and 19 pairs in hTF were added based on S–S proximity. Ca²⁺ and Cl[−] ions were added to a concentration of 100 mM, resulting in an initial size of 264,000 atoms. The system was equilibrated in stages for 13.5 ns, including 10 ns of fully unrestrained dynamics. The simulations were run using NAMD 2.7³³ in the NPT ensemble at a temperature of 310 K and a pressure of 1 atm; after the first 3.5 ns of equilibration, the area of the membrane was fixed. Other simulation parameters were set identically to those used previously⁴⁸. For steered molecular dynamics (SMD) simulations, the C α atom of the TbpA N-terminal plug domain residue was pulled in the $-z$ direction, away from the membrane, at a constant velocity of $5\ \text{\AA ns}^{-1}$ (refs 33, 49). To counterbalance the pulling forces, six residues at the extracellular periphery of the barrel domain were restrained in the z direction. An adaptive procedure was used to limit the maximum required system size during SMD simulations⁵⁰. When the extension of the unfolded region of the plug domain brought it near to the periodic boundary, the simulation was stopped, the unfolded region of the plug

domain distant from the barrel and membrane was cleaved, and the simulation restarted after a short equilibration of the water with the new N-terminal residue being pulled. With this procedure, used three times, approximately 150 Å of pulling was accomplished while keeping the system sizes below 300,000 atoms.

Sequence analysis and alignments. Sequence analysis and alignments were performed and analysed with the programs STRAP⁵¹ and JalView⁵². Figures were annotated and finalized with Adobe Illustrator.

34. Otwinowski, Z. & Minor, W. Processing of X-ray data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
35. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674 (2007).
36. Adams, P. D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **58**, 1948–1954 (2002).
37. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
38. Blanc, E. *et al.* Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta Crystallogr. D* **60**, 2210–2221 (2004).
39. Arnold, K., Bordoli, L., Kopp, J. & Schwede, T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* **22**, 195–201 (2006).
40. Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
41. Petoukhov, M. V. & Svergun, D. I. Analysis of X-ray and neutron scattering from biomacromolecular solutions. *Curr. Opin. Struct. Biol.* **17**, 562–571 (2007).
42. Svergun, O. & Genkina, O. A. The dependence of the dynamics of the extinction of a temporary connection on the recognizability of a reinforcing stimulus [in Russian]. *Zh. Vyssh. Nerv. Deiat. Im. I P Pavlova* **41**, 700–707 (1991).
43. Volkov, V. V. & Svergun, D. I. Uniqueness of ab initio shape determination in small-angle scattering. *J. Appl. Cryst.* **36**, 860–864 (2003).
44. Shaikh, T. R. *et al.* SPIDER image processing for single-particle reconstruction of biological macromolecules from electron micrographs. *Nature Protocols* **3**, 1941–1974 (2008).
45. Ludtke, S. J. 3-D structures of macromolecules using single-particle analysis in EMAN. *Methods Mol. Biol.* **673**, 157–173 (2010).
46. Heymann, J. B. & Belnap, D. M. Bsoft: image processing and molecular modeling for electron microscopy. *J. Struct. Biol.* **157**, 3–18 (2007).
47. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).
48. Gumbart, J., Wiener, M. C. & Tajkhorshid, E. Coupling of calcium and substrate binding through loop alignment in the outer-membrane transporter BtuB. *J. Mol. Biol.* **393**, 1129–1142 (2009).
49. Sotomayor, M. & Schulten, K. Single-molecule experiments *in vitro* and *in silico*. *Science* **316**, 1144–1148 (2007).
50. Gumbart, J., Wiener, M. C. & Tajkhorshid, E. Mechanics of force propagation in TonB-dependent outer membrane transport. *Biophys. J.* **93**, 496–504 (2007).
51. Gille, C. & Frommel, C. STRAP: editor for STRuctural Alignments of Proteins. *Bioinformatics* **17**, 377–378 (2001).
52. Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M. & Barton, G. J. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).

Clonal selection drives genetic divergence of metastatic medulloblastoma

Xiaochong Wu¹, Paul A. Northcott¹, Adrian Dubuc¹, Adam J. Dupuy², David J. H. Shih¹, Hendrik Witt³, Sidney Croul⁴, Eric Bouffet⁵, Daniel W. Fuhs⁶, Charles G. Eberhart⁷, Livia Garzia¹, Timothy Van Meter⁸, David Zagzag⁹, Nada Jabado¹⁰, Jeremy Schwartzentruber¹¹, Jacek Majewski¹⁰, Todd E. Scheetz², Stefan M. Pfister³, Andrey Korshunov¹², Xiao-Nan Li¹³, Stephen W. Scherer¹⁴, Yoon-Jae Cho¹⁵, Keiko Akagi¹⁶, Tobey J. MacDonald¹⁷, Jan Koster¹⁸, Martin G. McCabe¹⁹, Aaron L. Sarver²⁰, V. Peter Collins²¹, William A. Weiss²², David A. Largaespada²⁰, Lara S. Collier²³ & Michael D. Taylor²⁴

Medulloblastoma, the most common malignant paediatric brain tumour, arises in the cerebellum and disseminates through the cerebrospinal fluid in the leptomeningeal space to coat the brain and spinal cord¹. Dissemination, a marker of poor prognosis, is found in up to 40% of children at diagnosis and in most children at the time of recurrence. Affected children therefore are treated with radiation to the entire developing brain and spinal cord, followed by high-dose chemotherapy, with the ensuing deleterious effects on the developing nervous system². The mechanisms of dissemination through the cerebrospinal fluid are poorly studied, and medulloblastoma metastases have been assumed to be biologically similar to the primary tumour^{3,4}. Here we show that in both mouse and human medulloblastoma, the metastases from an individual are extremely similar to each other but are divergent from the matched primary tumour. Clonal genetic events in the metastases can be demonstrated in a restricted subclone of the primary tumour, suggesting that only rare cells within the primary tumour have the ability to metastasize. Failure to account for the bicompartamental nature of metastatic medulloblastoma could be a major barrier to the development of effective targeted therapies.

Thirty percent of patched-1-heterozygous (*Ptch*^{+/-}) mice develop non-disseminated medulloblastoma by 8 months of age⁵. Recently, the *Sleeping Beauty* (SB) transposon system was shown to be an effective tool for functional genomics studies of solid tumour initiation and progression^{6,7}. We expressed the SB11 transposase in cerebellar progenitor cells in transgenic mice under the *Math1* (also known as *Atoh1*) enhancer/promoter, but we did not observe any tumours when these mice were bred with mice transgenic for a concatemer of the T2/Onc transposon⁸ (Fig. 1a–j and Supplementary Figs 1 and 2). However, on a *Ptch*^{+/-} background, these *Math1-SB11/T2Onc* mice showed increased penetrance of medulloblastoma (~97%; 271 of 279 mice) compared with controls (~39%; 54 of 139 mice), as well as decreased latency (2.5 months compared with 8 months) (Fig. 1 and Supplementary Fig. 2). Although *Ptch*^{+/-} medulloblastomas are usually localized, the addition of SB transposition results in metastatic dissemination through the cerebrospinal fluid pathways, identical to the pattern that

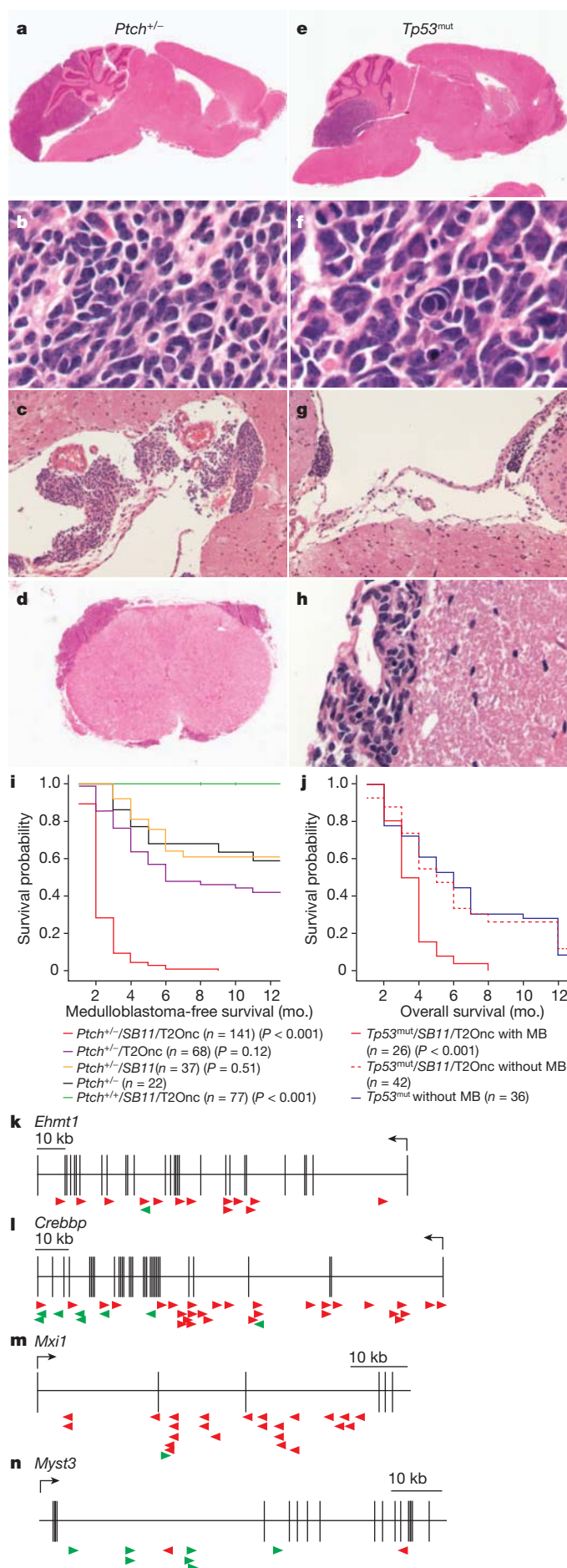
is seen in human children (Fisher's exact test, $P = 1.8 \times 10^{-7}$, odds ratio = 5.2; Supplementary Table 1) (Fig. 1c, d and Supplementary Fig. 2). As neither transposon nor transposase alone had an effect on tumour incidence, latency or dissemination, we conclude that SB-induced insertional mutagenesis drives medulloblastoma progression on the *Ptch*^{+/-} background (Fig. 1i and Supplementary Fig. 2).

Humans with germline mutations in the tumour-suppressor gene *TP53* have Li-Fraumeni syndrome and have an increased risk of developing medulloblastoma. Although no medulloblastomas were found in mice with mutant *Tp53* (also known as *Trp53*) (denoted *Tp53*^{mut} mice, which includes *Tp53*^{+/-} and *Tp53*^{-/-}), 40% of *Tp53*^{mut}/*Math1-SB11/T2Onc* mice developed disseminated medulloblastoma⁹ (Fig. 1e–h, j and Supplementary Fig. 2). Human medulloblastomas with *TP53* mutations frequently have large cell/anaplastic histology. *Tp53*^{mut}/*Math1-SB11/T2Onc* medulloblastomas have large cells, nuclear atypia and nuclear moulding that is typical of large cell/anaplastic histology (Fig. 1f). We conclude that SB transposition can drive the initiation and progression of metastatic medulloblastoma on a *Tp53*^{mut} background.

We used linker-mediated PCR and 454 sequencing to identify the site of T2/Onc insertions in *Ptch*^{+/-}/*Math1-SB11/T2Onc* and *Tp53*^{mut}/*Math1-SB11/T2Onc* primary medulloblastomas and their matched metastases. Genes that contained insertions statistically more frequently than the background rate were identified as gene-centric common insertion sites (gCISs)¹⁰. We identified 359 gCISs in 139 primary tumours on the *Ptch*^{+/-} background and 26 gCISs in 36 primary medulloblastomas on the *Tp53*^{mut} background (Supplementary Tables 2–7 and Supplementary Figs 3–5). A large number of gCISs were candidate medulloblastoma oncogenes or tumour-suppressor genes¹¹ (Supplementary Table 8). Insertions in candidate tumour-suppressor genes, including *Ehmt1*, *Crebbp* and *Mxi1*, are predicted to cause a loss of function (Fig. 1k–m), whereas insertions in putative medulloblastoma oncogenes are largely gain of function, as exemplified by *Myst3* (Fig. 1n).

Many gCISs mapped to regions of amplification, focal hemizygous deletion and homozygous deletion (Supplementary Table 8) that we

¹Arthur and Sonia Labatt Brain Tumour Research Center, and Program in Developmental and Stem Cell Biology, The Hospital for Sick Children, Toronto, Ontario M5G 1X8, Canada. ²Molecular & Cellular Biology Program, The University of Iowa, Iowa City, Iowa 52242, USA. ³German Cancer Research Center (DKFZ), and Department of Pediatric Oncology, Hematology and Immunology, University Hospital Heidelberg, Im Neuenheimer Feld 280, 69120 Heidelberg, Germany. ⁴University Health Network Pathology, Arthur and Sonia Labatt Brain Tumour Research Centre, Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario M5S 1A1, Canada. ⁵Brain Tumour Program, Haematology and Oncology, The Hospital for Sick Children, Toronto, Ontario M5G 1X8, Canada. ⁶Department of Neurosurgery, University of Utah School of Medicine, Salt Lake City, Utah 84132, USA. ⁷Johns Hopkins University, Baltimore, Maryland 21210, USA. ⁸Virginia Commonwealth University, Richmond, Virginia 23284, USA. ⁹Department of Pathology and Neurosurgery, Division of Neuropathology, New York University School of Medicine, New York, New York 10016, USA. ¹⁰Departments of Human Genetics and Experimental Medicine, McGill University, Montreal H3Z 2Z3, Quebec, Canada. ¹¹McGill University and Genome Quebec Innovation Centre, Montreal, Quebec H3A 0G1, Canada. ¹²German Cancer Research Institute (DKFZ) and Department of Neuropathology, University of Heidelberg, Im Neuenheimer Feld 220/221, 69120 Heidelberg, Germany. ¹³Brain Tumour Program, Texas Children's Cancer Center, and Department of Pediatrics, Baylor College of Medicine, Houston, Texas 77030, USA. ¹⁴Program in Genetics and Genomic Biology and The Centre for Applied Genomics, The Hospital for Sick Children, and McLaughlin Centre and Department of Molecular Genetics, University of Toronto, Toronto, Ontario M5S 1A1, Canada. ¹⁵Departments of Neurology and Neurosurgery, Stanford University School of Medicine, Stanford, California 94305, USA. ¹⁶Department of Molecular Virology, Immunology and Medical Genetics, The Ohio State University Comprehensive Cancer Center, Columbus, Ohio 43210, USA. ¹⁷Pediatric Neuro-Oncology Program, Emory University School of Medicine, Atlanta, Georgia 30307, USA. ¹⁸Department of Human Genetics, Academic Medical Center, University of Amsterdam, 1100 DE Amsterdam, The Netherlands. ¹⁹School of Cancer and Enabling Sciences, University of Manchester, Manchester M20 4BX, UK. ²⁰Masonic Cancer Center, University of Minnesota, Minneapolis, Minnesota 55455, USA. ²¹Department of Pathology, University of Cambridge, Cambridge CB2 1QP, UK. ²²Departments of Neurology, Pediatrics and Neurological Surgery, University of California, San Francisco, California 94143, USA. ²³School of Pharmacology, University of Wisconsin, Madison, Wisconsin 53715, USA. ²⁴Division of Neurosurgery, Arthur and Sonia Labatt Brain Tumour Research Center, and Program in Developmental and Stem Cell Biology, The Hospital for Sick Children, University of Toronto, Toronto, Ontario M5G 1X8, Canada.



recently reported in the genome of a large cohort of human medulloblastomas¹¹. There is a high level of overlap between gCISs and known cancer genes (in the Catalogue of Somatic Mutations in Cancer (COSMIC) database) (Supplementary Tables 9 and 10), suggesting that many gCISs are bona fide driver genes in medulloblastoma (Fisher's exact test, *P* = 0.0012)¹². Similarly, many of the mouse gCISs and the genes amplified in human medulloblastomas are over-expressed in human *SHH*-driven medulloblastomas (Supplementary Fig. 6). Conversely, mouse gCISs hemizygously deleted in human medulloblastomas were frequently expressed at a lower level in human medulloblastomas (Supplementary Fig. 6). The expression of six out of seven gCISs that had been studied by immunohistochemistry on a human medulloblastoma tissue microarray was associated with significantly worse overall and progression-free survival in human medulloblastoma¹³ (Supplementary Table 11 and Supplementary Figs 7 and 8). We conclude that our SB-driven leptomeningeal-disseminated medulloblastoma model resembles the human disease anatomically, pathologically and genetically and thus is an accurate model of the human disease that can be used to identify candidate driver events and understand the pathogenesis of human medulloblastoma.

We compared the gCISs identified from *Ptch*^{+/-}/*Math1*-SB11/T2Onc and *Tp53*^{mut}/*Math1*-SB11/T2Onc primary medulloblastomas and matched metastases (Supplementary Table 2). Strikingly, the overlap between primary tumour gCISs (pri-gCISs) from *Ptch*^{+/-}/*Math1*-SB11/T2Onc tumours and those from metastases (met-gCISs) from the same animals was only 9.3% of all gCISs (Fig. 2a). Similarly, the overlap between pri-gCISs from *Tp53*^{mut}/*Math1*-SB11/T2Onc mice and the matching met-gCISs was only 8.9% (Fig. 2b). The leptomeningeal metastases and the matched primary tumour have identical, highly clonal insertion sites on both genetic backgrounds (Fig. 2c). The probability of two (or three) unrelated tumours having SB insertions in exactly the same TA dinucleotide is extremely low. We conclude that the leptomeningeal metastases and the matched primary tumour arise from a common transformed progenitor cell and have subsequently undergone genetic divergence.

Sequencing also identified insertions that are highly clonal in the metastases but are not observed in the matched primary tumour (data not shown). End-point PCR for these insertions in the matched primary and metastatic tumours shows that the insertion is highly clonal in the metastasis (or metastases) and is present in a very small subclone of the primary tumour (Fig. 2d and Supplementary Fig. 9). These data are consistent with a model in which metastatic disease arises from a minor restricted subclone of the primary tumour. Dissemination could occur repeatedly from the same subclone of the primary tumour, which seeds the rest of the central nervous system, or it could occur once, followed by reseeding of the rest of the leptomeningeal space by the initial metastasis. Insertions that are restricted to a minor subclone of the

Figure 1 | Transposon mutagenesis models of disseminated human medulloblastoma. **a–d**, The histology of transposon-driven medulloblastoma on the *Ptch*^{+/-} background resembles human medulloblastoma, with leptomeningeal metastases on the surface of the brain (**c**) and spinal cord (**d**). Images show haematoxylin and eosin staining (**a**, entire brain; **b**, upper spinal cord). **e–h**, The histology of transposon-driven medulloblastoma on the *Tp53*^{mut} background shows histological features of large cell/anaplastic medulloblastoma, including nuclear pleomorphism and nuclear wrapping (**f**). Dissemination to the leptomeningeal spaces of the brain (**g**) and spinal cord (**h**) also occurs on this background. **i**, *Ptch*^{+/-} mice with SB transposition develop more frequent medulloblastomas with a shorter latency than *Ptch*^{+/-} mice without transposition. *P* values are from *t*-tests of survival comparing individual genotypes to *Ptch*^{+/-} mice; *n*, number of mice per genotype. mo., months. **j**, Medulloblastoma (MB) was not observed in *Tp53*^{mut} mice without transposition but was observed in 42% of *Tp53*^{mut} mice with transposition. *P* values are from *t*-tests comparing survival between *Tp53*^{mut} mice and *Tp53*^{mut}/SB11/T2Onc mice with MB; *n*, number of mice. **k–n**, Insertion maps of notable gCISs. Insertions in the direction of transcription are denoted by green arrows, and those against the direction of transcription are denoted by red arrows. Transcription start sites are denoted by black arrows.

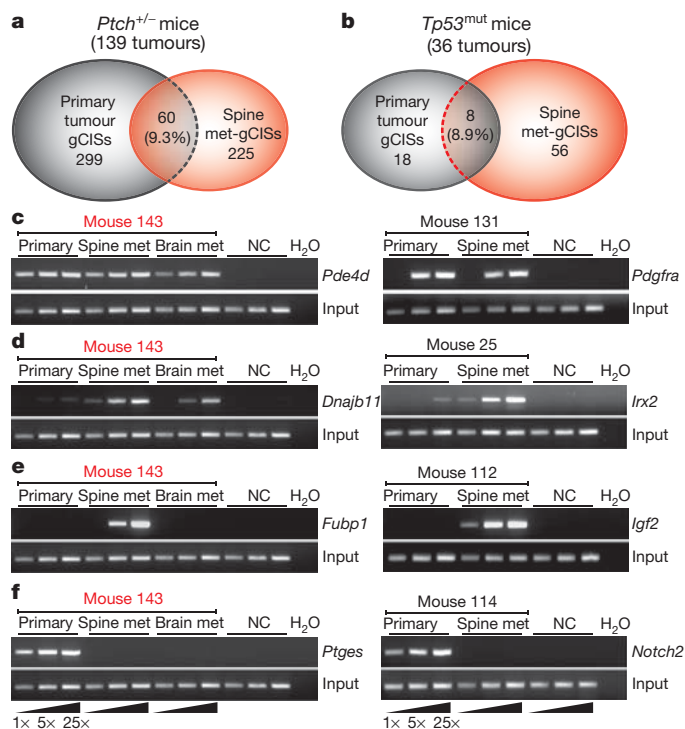


Figure 2 | Transposon-driven metastatic medulloblastoma genetically differs from the primary tumour. **a, b**, Venn diagrams depicting the degree of overlap and discordance in the gCISs in primary tumours and metastases, on the *Ptch*^{+/-} and *Tp53*^{mut} backgrounds. **c–f**, Insertion-site end-point PCR was used to demonstrate the relative clonality of insertions between samples. Data for medulloblastoma in five mice are shown (mouse 143, left; and four mice, right). Three levels of input DNA were used for each sample (1×, 5× and 25×, with the increase depicted by a wedge). Shown are clonal events found in both the primary tumour and matching metastases (met) (**c**), insertions that are highly clonal in the metastases but very subclonal in the matching primary tumour (**d**), insertions that are highly clonal in the metastases but undetectable in the matching primary tumour (**e**), and insertions that are highly clonal in the primary tumour but undetectable in the matching metastases (**f**). NC, negative control; genomic DNA from a *Math1-SB11/T2Onc* double-transgenic mouse cerebellum.

primary tumour but that are clonal in the metastases could correspond to the previously described ‘metastasis virulence’ genes¹⁴, which offer a genetic advantage during dissemination but not to the primary tumour. Another explanation for our data could be that the primary tumour was reseeded by a metastatic clone that had acquired additional genetic events in the periphery. This hypothesis is mitigated by the presence of highly clonal insertions in the metastasis that are completely absent from the primary tumour in the same animal¹⁵ (Fig. 2e). As reseeding should be accompanied by contamination of the primary tumour with events found in the metastases, the absence of these events in the matched primary tumour makes reseeding much less likely (Fig. 2e). We propose that events found in only one metastasis represent progression events that are acquired post metastasis and that could lead to localized progression of metastatic disease, as is sometimes observed in human children.

We observed highly clonal insertions in the primary tumour, including in known medulloblastoma oncogenes such as *Notch2* and *Tert*, that were not found in the matching metastases (Fig. 2f). This pattern could be explained through remobilization of the SB transposon in the metastatic tumour; however, no signs of the DNA footprint remaining after SB remobilization at these loci were observed¹⁶ (Supplementary Fig. 10). We suggest that these events, which may constitute driver events in the primary tumour, have arisen in the primary tumour after the metastases have disseminated (post-dispersion events). Although these known oncogenes are attractive targets for therapy,

their utility as targets may be limited if the targets are not also found in the leptomeningeal compartment of the disease. Our data from two separate mouse lines support a model in which medulloblastoma disseminates early from a restricted subclone of the primary tumour and in which the primary tumour and the matched metastases then undergo differential clonal selection and evolution. Failure to account for the differences between the primary and leptomeningeal compartments could lead to the failure of targeted therapies. Failure to study the leptomeningeal disease (Fig. 2d, e) could result in systematically overlooking crucial targets for therapy in this compartment.

Examining the met-gCISs using Gene Set Enrichment Analysis (GSEA) demonstrated differences between the primary and metastatic disease, including enrichment for genes involved in the cytoskeleton in metastases (Supplementary Table 12). Targets that are present in both compartments and are maintenance genes, as exemplified by *Pdgfra*, will be optimal targets for treating both the primary tumour and the metastases (Fig. 2c and Supplementary Tables 7 and 9).

Pten, *Akt2*, *Igf2* and *Pik3r1* are all met-gCISs, implicating the phosphatidylinositol-3-OH kinase (PI(3)K) pathway in medulloblastoma progression. We injected the cerebellum of Nestin-TVA mice¹⁷ with either an *Shh*-overexpressing retroviral vector (denoted *Shh* virus) or an *Shh*- and *Akt*-overexpressing retroviral vector (denoted *Shh* + *Akt* virus). Cerebellar injection of *Shh* virus alone resulted in medulloblastomas in 6 of 41 animals, compared with 20 of 42 animals injected with *Shh* + *Akt* virus ($P = 0.0018$). Although metastases were not observed with *Shh* virus alone (0 of 41), medulloblastoma metastases were observed in 9 of 42 animals injected with *Shh* + *Akt* virus ($P = 0.0024$) (Supplementary Fig. 11). *In vivo* modelling validates PI(3)K signalling as a putative contributor to leptomeningeal dissemination of medulloblastoma.

Previous publications and clinical approaches to human medulloblastoma have largely assumed that the primary tumour and its matched metastases are highly similar^{3,4}. To test this assertion, we formally reviewed all cases of medulloblastoma from the past decade at The Hospital for Sick Children, in Ontario, Canada, and we identified 19 patients who had bulk residual primary tumour after surgery and metastases visible by magnetic resonance imaging, both of which could be followed for response to treatment (Supplementary Fig. 12 and Supplementary Table 13). Although it is possible that the metastases received less radiotherapy than the primary tumour in a subset of patients, in 58% of all cases (11 of 19) we observed a disparate response to therapy between the primary tumour and the matched metastases (binomial test, $P < 2.2 \times 10^{-16}$). Identification of definitive differences in the clinical response to standard therapy between the primary and the metastatic compartment awaits the completion of large, well-controlled, prospective clinical trials.

We examined seven matched primary and metastatic medulloblastomas for copy number aberrations (Fig. 3, Supplementary Figs 13 and 14, and Supplementary Tables 14 and 15). In each case, the primary tumour and the matched metastases shared complicated genetic events that provide strong support for their descent from a common transformed progenitor cell. Similar to our mouse data, in each case we observed clonal genetic events in the metastatic tumour(s) that were not present in the matched primary tumour (Fig. 3 and Supplementary Fig. 14). We also observed genetic events in the primary tumour that were absent from the matched metastases, consistent with a post-dispersion event (Fig. 3 and Supplementary Fig. 14). One patient with multiple leptomeningeal metastases had a deletion of chromosome 1p in only one of three examined metastases (Fig. 3a). This pattern of genetic events being present in only a subset of metastases could be a mechanism for the emergence of therapy-resistant metastatic clones.

We performed interphase fluorescence *in situ* hybridization (FISH) for the known medulloblastoma oncogenes *MYCN* and *MYC* on a collection of 17 paraffin-embedded primary and metastatic pairs of human medulloblastomas^{18–20}. *MYCN* was amplified in three primary medulloblastomas but not in the matching metastases (Fig. 3b and

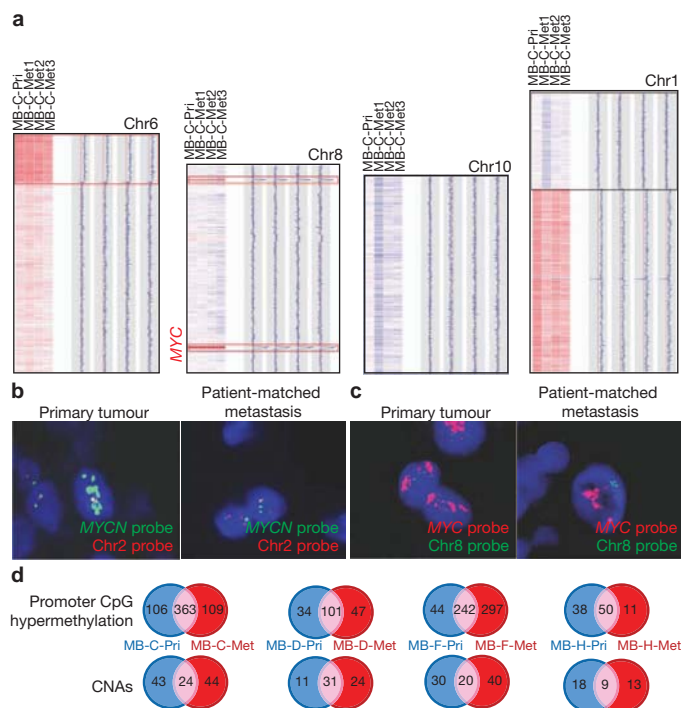


Figure 3 | Human medulloblastoma metastases are biologically distinct from their matched primary tumour. **a**, Copy number data from a primary medulloblastoma (MB-C-Pri) and three patient-matched metastases (MB-C-Met1, MB-C-Met2 and MB-C-Met3), with chromosomal regions in red representing genetic gain (amplification) and in blue denoting genetic loss (deletion). Examples of shared clonal events (red boxes) and events limited to one but not all metastases (black box) are shown. Chr, chromosome. **b**, Interphase FISH shows amplification of *MYCN* in a primary tumour but not the matched metastasis. Nuclei appear blue owing to 4',6-diamidino-2-phenylindole (DAPI) staining. **c**, Interphase FISH for *MYC* demonstrates amplification in both the primary tumour and its matched metastases. **d**, Venn diagrams depicting the degree of overlap and discordance in promoter CpG methylation events and CNAs in primary medulloblastomas and their matched metastases, with MB-C, MB-D and MB-F and MB-H denoting different patients.

Supplementary Fig. 15). Conversely, *MYC* was amplified in two primary tumours and their matching metastases (Fig. 3c). These data are consistent with *MYCN* amplification being a post-dispersion event, similar to examples in SB-driven mouse medulloblastoma, and strongly indicate that anti-*MYCN* therapeutics may lack efficacy in the metastatic compartment of human medulloblastoma. The possibility that *MYCN* amplicons in the metastases have been 'lost' over time cannot be excluded.

We subsequently analysed promoter CpG methylation in these matched pairs and found much discordance between the primary tumour and matched metastases (Fig. 3d, Supplementary Figs 13 and 16, and Supplementary Tables 16 and 17). Finally, we performed whole-exome sequencing on a limited set of matched primary and metastatic medulloblastomas and found many single nucleotide variants (SNVs) that were restricted to a single compartment (Supplementary Fig. 13 and Supplementary Table 18). The discordance of CNAs, promoter CpG methylation events and SNVs between the primary tumour and its matched metastases supports a bicompartmental model for metastatic medulloblastoma. The mutational load in the human tumours (the combination of CNAs, CpG methylation and SNVs) compares favourably with the mutational load in our transposon-driven mouse models (in which the median number of gCISs is 25 per tumour; Supplementary Table 19). Validation of the individual CNAs that were restricted to the metastases showed that these CNAs can be detected in a very minor subclone of the primary tumour, in keeping with the relationship identified in the mouse model (Supplementary Fig. 17 and Supplementary Tables 20 and 21). Pathway analysis using

the Database for Annotation, Visualization and Integrated Discovery (DAVID) to compare mouse gCISs with the genes that were affected in the human metastases identified only one statistically significant shared signalling pathway: insulin signalling ($P = 0.027$) (Supplementary Table 22). The known role of insulin receptor signalling in primary medulloblastoma²¹, together with the data presented here on the role of AKT in metastatic medulloblastoma, suggests that insulin signalling should be prioritized as a therapeutic target to be tested in clinical trials.

We performed unsupervised hierarchical clustering on the CpG methylation data, and we found that normal cerebellar controls cluster away from the medulloblastomas, whereas metastases cluster with their matching primary tumour (Fig. 4a). However, metastases cluster

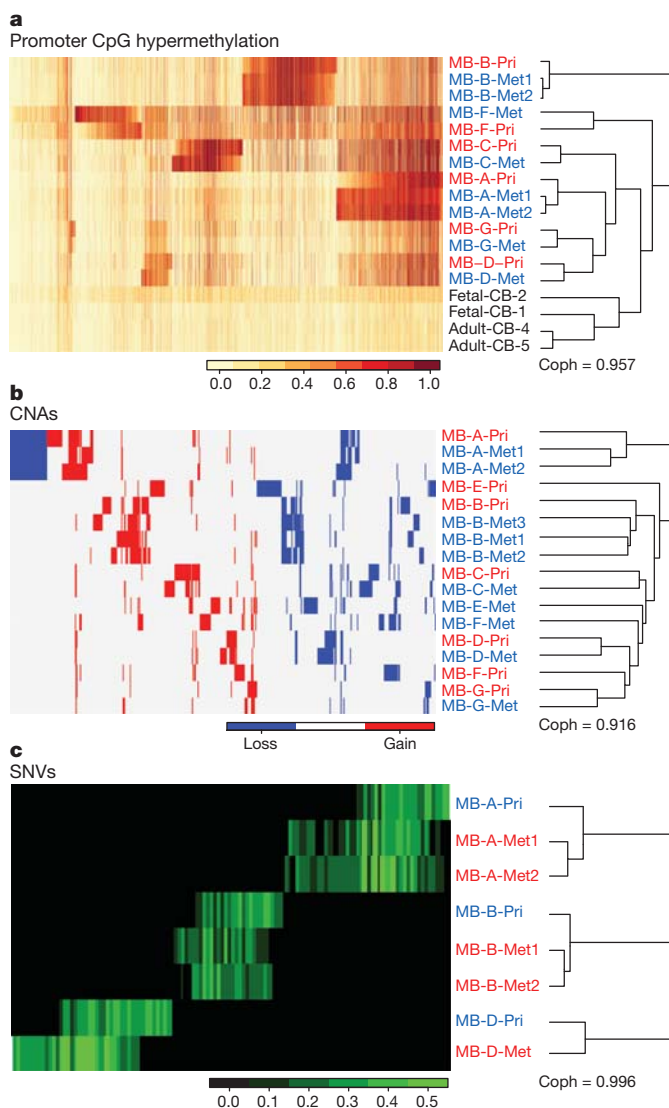


Figure 4 | Human medulloblastoma metastases are genetically distinct from their matched primary tumour. **a**, Profiling the methylation status of 27,578 CpG dinucleotide sites in the human genome in a collection of human matched primary and metastatic medulloblastomas; the top 2,000 genes are shown. Unsupervised hierarchical clustering by CpG methylation pattern demonstrates that patient-matched metastases are more similar to each other than to the matched primary tumour. **b**, Unsupervised clustering of regions of copy number gain and loss demonstrates that patient-matched metastases are more similar to each other than to the matched primary tumour. **c**, Unsupervised hierarchical clustering of SNV data from whole-exome sequencing demonstrates that patient-matched metastases are more similar to each other than to the matched primary tumour. SNVs that are found only in the primary compartment or only in both examined tumours in the metastatic compartment are evident. Coph, cophenetic correlation coefficient.

closer to each other than they do to the matched primary tumour (z -test, $P = 0.0014$) (Supplementary Fig. 18). Unsupervised hierarchical clustering of CNA and exome SNV data uncovered the same relationships (Fig. 4b, c). Evident within the exome data are many events that are shared only by patient-matched metastases (that is, metastases from a single patient), as well as events that are restricted to the primary tumour, both of which are similar to the genetic patterns observed in mice. These three data sets support a model in which patient-matched human medulloblastoma metastases are epigenetically and genetically very similar to each other but have substantially diverged from the primary tumour, resulting in two different disease compartments: the primary and metastatic compartments.

Our data from two mouse models, with support from initial data from human medulloblastoma, suggest that leptomeningeal metastases of medulloblastoma from a single human or mouse are genetically similar to each other but are highly divergent from the matched primary tumour, consistent with a bicompartamental model of disease. Our results are consistent with a model in which metastases arise from a restricted subclone of the primary tumour through a process of clonal selection in both humans and mice. That metastases might arise from a pre-existing minor subclone of the primary tumour through clonal selection was suggested more than three decades ago, but it remains a controversial hypothesis that might not be true of all cancers^{22–25}. Failure to account for the divergent molecular pathology of the metastatic compartment may result in selection of therapeutic targets present in the primary tumour, which is more amenable to surgical control, but not the metastases, which are the more frequent cause of death.

METHODS SUMMARY

Generation of *Math1-SB11* construct. *SB11* cDNA was excised from the vector pCMV-*SB11* and ligated into the vector J2Q-*Math1* (refs 8, 26).

Linker-mediated PCR and 454 deep sequencing. Bar-coded, linker-mediated PCR was performed as previously described⁶. Sample preparation for the 454 sequencing and the subsequent procedures was performed as previously described²⁷.

Determination of gCISs. A chi-squared analysis was performed to determine whether the number of observed integration events within each transcription unit in the SB-driven medulloblastomas was significantly greater than expected given the following: the number of TA dinucleotide sites within the gene relative to the number of TA sites in the genome, the number of integration sites within each tumour, and the total number of tumours in each cohort. This gCIS analysis produced a P value for each of the ~19,000 mouse RefSeq genes, and Bonferroni correction was therefore used to adjust for multiple hypothesis testing.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 10 October 2010; accepted 3 January 2012.

Published online 15 February 2012.

- Gajjar, A. *et al.* Risk-adapted craniospinal radiotherapy followed by high-dose chemotherapy and stem-cell rescue in children with newly diagnosed medulloblastoma (St Jude Medulloblastoma-96): long-term results from a prospective, multicentre trial. *Lancet Oncol.* **7**, 813–820 (2006).
- Mabbott, D. J. *et al.* Serial evaluation of academic and behavioral outcome after treatment with cranial radiation in childhood. *J. Clin. Oncol.* **23**, 2256–2263 (2005).
- MacDonald, T. J. *et al.* Expression profiling of medulloblastoma: PDGFRA and the RAS/MAPK pathway as therapeutic targets for metastatic disease. *Nature Genet.* **29**, 143–152 (2001).
- Ramaswamy, S., Ross, K. N., Lander, E. S. & Golub, T. R. A molecular signature of metastasis in primary solid tumors. *Nature Genet.* **33**, 49–54 (2003).
- Goodrich, L. V., Milenkovic, L., Higgins, K. M. & Scott, M. P. Altered neural cell fates and medulloblastoma in mouse patched mutants. *Science* **277**, 1109–1113 (1997).
- Collier, L. S., Carlson, C. M., Ravimohan, S., Dupuy, A. J. & Largaespada, D. A. Cancer gene discovery in solid tumours using transposon-based somatic mutagenesis in the mouse. *Nature* **436**, 272–276 (2005).
- Dupuy, A. J., Akagi, K., Largaespada, D. A., Copeland, N. G. & Jenkins, N. A. Mammalian mutagenesis using a highly mobile somatic *Sleeping Beauty* transposon system. *Nature* **436**, 221–226 (2005).

- Helms, A. W., Abney, A. L., Ben-Arie, N., Zoghbi, H. Y. & Johnson, J. E. Autoregulation and multiple enhancers control *Math1* expression in the developing nervous system. *Development* **127**, 1185–1196 (2000).
- Wetmore, C., Eberhart, D. E. & Curran, T. Loss of *p53* but not *ARF* accelerates medulloblastoma in mice heterozygous for *patched*. *Cancer Res.* **61**, 513–516 (2001).
- Brett, B. T. *et al.* Novel molecular and computational methods improve the accuracy of insertion site analysis in *Sleeping Beauty*-induced tumors. *PLoS ONE* **6**, e24668 (2011).
- Northcott, P. A. *et al.* Multiple recurrent genetic events converge on control of histone lysine methylation in medulloblastoma. *Nature Genet.* **41**, 465–472 (2009).
- Forbes, S. A. *et al.* in *Current Protocols in Human Genetics* Ch. 10.11 doi:10.1002/0471142905.hg1011s57 (2008).
- Northcott, P. A. *et al.* Medulloblastoma comprises four distinct molecular variants. *J. Clin. Oncol.* **29**, 1408–1414 (2011).
- Nguyen, D. X. & Massague, J. Genetic determinants of cancer metastasis. *Nature Rev. Genet.* **8**, 341–352 (2007).
- Norton, L. & Massague, J. Is cancer a disease of self-seeding? *Nature Med.* **12**, 875–878 (2006).
- Luo, G., Ivics, Z., Izsvák, Z. & Bradley, A. Chromosomal transposition of a Tc1/mariner-like element in mouse embryonic stem cells. *Proc. Natl Acad. Sci. USA* **95**, 10769–10773 (1998).
- Rao, G., Pedone, C. A., Coffin, C. M., Holland, E. C. & Fuhs, D. W. c-Myc enhances sonic hedgehog-induced medulloblastoma formation from nestin-expressing neural progenitors in mice. *Neoplasia* **5**, 198–204 (2003).
- Swartling, F. J. *et al.* Pleiotropic role for *MYCN* in medulloblastoma. *Genes Dev.* **24**, 1059–1072 (2010).
- Pfister, S. *et al.* Outcome prediction in pediatric medulloblastoma based on DNA copy-number aberrations of chromosomes 6q and 17q and the *MYC* and *MYCN* loci. *J. Clin. Oncol.* **27**, 1627–1636 (2009).
- Korshunov, A. *et al.* Accumulation of genomic aberrations during clinical progression of medulloblastoma. *Acta Neuropathol.* **116**, 383–390 (2008).
- Hahn, H. *et al.* *Patched* target *Igf2* is indispensable for the formation of medulloblastoma and rhabdomyosarcoma. *J. Biol. Chem.* **275**, 28341–28344 (2000).
- Fidler, I. J. & Kripke, M. L. Metastasis results from preexisting variant cells within a malignant tumor. *Science* **197**, 893–895 (1977).
- Scheel, C., Onder, T., Karnoub, A. & Weinberg, R. A. Adaptation versus selection: the origins of metastatic behavior. *Cancer Res.* **67**, 11476–11480 (2007).
- Talmadge, J. E. Clonal selection of metastasis within the life history of a tumor. *Cancer Res.* **67**, 11471–11475 (2007).
- Talmadge, J. E. & Fidler, I. J. The biology of cancer metastasis: historical perspective. *Cancer Res.* **70**, 5649–5669 (2010).
- Geurts, A. M. *et al.* Gene transfer into genomes of human cells by the *Sleeping Beauty* transposon system. *Mol. Ther.* **8**, 108–117 (2003).
- Starr, T. K. *et al.* A transposon-based genetic screen in mice identifies genes altered in colorectal cancer. *Science* **323**, 1747–1750 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements M.D.T. holds a Canadian Institutes of Health Research Clinician-Scientist Phase II Award, was a Sontag Foundation Distinguished Scholar, and is supported by grants from the National Institutes of Health (R01CA148699), the Pediatric Brain Tumor Foundation, the Canadian Cancer Society, and Brainchild. X.W. was supported by a fellowship from the American Brain Tumor Association in tribute to Tracy Greenwood. L.G. was supported by a fellowship from the Davis M. Ferguson Fund from the American Brain Tumor Association. A.D. was supported by a Vanier Doctoral Fellowship from the Canadian Institutes of Health Research. L.S.C. was supported by a grant (K01CA122183) and a Kimmel Scholar award from the Kimmel Foundation. C.E. was supported by a grant from the National Institutes of Health (NS055089). We thank S. Archer for technical writing assistance.

Author Contributions M.D.T., X.W., P.A.N., L.S.C., A.D. and D.L. conceived the research and planned the experiments. X.W., P.A.N., A.D., L.G. and D.J.H.S. carried out the vast majority of the experiments under M.D.T.'s guidance. C.E., T.V.M., D.Z., Y.-J.C., T.M., X.-N.L., V.P.C., M.G.M. and W.A.W. provided human tumour materials. All authors contributed experimental expertise and participated in the writing of the manuscript. A.J.D., D.J.H.S., T.E.S., S.W.S., K.A., J.K., A.L.S., D.L. and L.S.C. provided biostatistical and bioinformatic expertise. E.B. provided the clinical data and analysis. D.W.F. carried out the *Akt* experiments. N.J., J.S. and J.M. carried out the exome sequencing. H.W., S.M.P. and A.K. carried out the immunostaining of human medulloblastoma tissue microarrays and FISH for *MYCN* and *MYC*. S.C. carried out the pathological analysis of mouse brain tumours.

Author Information CpG methylation data have been deposited in the Gene Expression Omnibus under accession number GSE34356. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.D.T. (mdtaylor@sickkids.ca).

METHODS

Linker-mediated PCR and 454 deep sequencing. Genomic DNA was isolated and purified from mouse tissues with a DNeasy Blood & Tissue Kit (QIAGEN). The subsequent bar-coded, linker-mediated PCR was performed as previously described⁶. Sample preparation for the 454 sequencing and the subsequent procedures was performed as previously described²⁷.

PCR for SB-tagged fragments. The primers for amplifying SB-transposon insertion sites were designed based on the chromosomal location of each independent insertion site and its orientation to transcription. The primers at the inverted repeats/direct repeats (left) (IRDRL) and inverted repeats/direct repeats (right) (IRDRR) of the transposon were 5'-CTGGGAATGTGATGAAAGAAATAAAA-3' and 5'-TTGTGTCATGCACAAAGTAGATGT-3', respectively. The input represents genomic DNA with SB transposition, which was illustrated by SB excision PCR that detected the transposon post transposition⁶. Three points of input (1×, 5× and 25×) were used. The following primers were used: *Pde4d*-143L, 5'-CACATAAAACTGGACACCTAG-3'; *Pdgfra*-131R, 5'-CTATCATGACCA CACGAAGAGAGTGAAC-3'; *Dnajb11*-143L, 5'-CATGAGCTATGGCACA GATAC-3'; *Fubp1*-143R, 5'-CACTAGTGCCCATGGATTAGG-3'; *Ptges*-143R, 5'-CAGAACTGATAGAGGCCAAAG-3'; *Irx2*-25L, 5'-CAACACTTTCAGAC ACACATATATC-3'; *Igf2*-112R, 5'-GTGACCAAGTGTGATTTCGTGGAATTT TTTGGG-3'; and *Notch2*-114R, 5'-CAGTGTCCAGGCAGTCATTCAAAGA GTG-3'. Details about the primer design for specific insertion sites and the PCR protocol are available on request.

Review of clinical cases. We systematically reviewed all cases of medulloblastoma seen at The Hospital for Sick Children (Toronto, Ontario) over the past ten years. Cases that have both metastases and post-operative residual bulky disease at the primary site were identified on the basis of post-operative imaging obtained within 72 h of surgery. All radiology results were reviewed by a senior neuro-oncologist (E.B.). Objective responses of both the primary tumour and the metastatic disease were measured using the standard International Society of Paediatric Oncology (SIOP) criteria for clinical trials of paediatric brain tumours²⁸.

End-point PCR on human samples. For PCRs to confirm the deletion of the *CDKN2A* locus on chromosome 9, a genome-walking approach (GenomeWalker Universal Kit, Clontech, Catalogue number 638904) was taken to locate the specific deletion region based on single nucleotide polymorphism (SNP) coordinates. The following primers flanking the deletion region were used: forward, 5'-GCAATTAACCAAGACCACCAATGGCAAG-3'; and reverse, 5'-GTAGC TATTGGGGAGGTTGAGAAGGAG-3'. Three points of input shown as *ACTB* (1×, 5× and 25×) were used. The PCR products were inserted into the pCR2.1 TA cloning vector (Invitrogen), sequenced and searched against the human genome in the blast database to confirm the deletion. For *REXOIL1* deletion on chromosome 8, specific primers flanking the deletion region were designed based on SNP microarray results. The PCR products were TA-cloned and sequenced as described above. The following primers were used: forward, 5'-GGCTGACTC CCTTCTGATATAG-3'; and reverse, 5'-CAATCACTTACAGTTACTAGGC AC-3'. Details about the primer design and PCR protocols are available on request.

Chromosomal mapping of gCISs. Chromosomal maps of gCIS-associated genes were obtained from the UCSC Mouse Genome Browser (assembly in July 2007). Each insertion site of a specific CIS was mapped to the gene with the same orientation as the direction of transcription (arrow in green) or the inverse orientation to the direction of transcription (arrow in red).

Human medulloblastoma tumour specimens. All tumour specimens were obtained in accordance with the Research Ethics Board at The Hospital for Sick Children. Surgically resected, fresh frozen samples were obtained from the Cooperative Human Tissue Network and the Brain Tumor Tissue Bank.

SB remobilization. Potential SB insertion sites at *Fubp1*, *Mnat1* or *Igf2* in primary tumours from mouse numbers 143, 14 or 11 or sites at *Ptges*, *Aof1* and *Notch2* in the matched spine metastases were tested for remobilization. The primers were designed to amplify each insertion site to produce approximately 300 base pairs (bp) with the insertion site in the middle. PCR products were either sequenced directly or after being TA-cloned. The resultant sequences were examined for 'scars' from potential remobilization. As positive controls for the scars, primers were used to amplify the T2/Onc transposon in each sample⁶. The products were sequenced and examined for the scars as described above. The following primers were used: *Aof1* forward (Fw), 5'-TACTCCAGACAGTCAGTCAGTG-3'; *Aof1* reverse (Rv), 5'-TAGTTCTGCCTCATGCCACAAG-3'; *Ptges* Fw, 5'-ACAGAG AAGGCTTCAGAGCTC-3'; *Ptges* Rv, 5'-GGTGCTCTCTGCTGTCCAATC-3'; *Notch2* Fw, 5'-CAAGCTTTCAAGTATAAACCACGC-3'; *Notch2* Rv, 5'-GAAT GCATCATCCAGTGTCCAG-3'; *Fubp1* Fw, 5'-AGGAACGGGCTGGTGTAA AATG-3'; *Fubp1* Rv, 5'-TCTAATACCAATTCTTGGCTTGC-3'; *Mnat1* Fw, 5'-CTAACACATCAGAGTTGGACAAG-3'; *Mnat1* Rv, 5'-CATGAAGACCTG AGAGTGCAG-3'; *Igf2* Fw, 5'-GTGATTGGTGAATGTACTCTTTCC-3'; and

Igf2 Rv, 5'-GTGGAACACTAGATTCTGTAGTC-3'. Details about the primer design and PCR protocols are available on request.

Hierarchical clustering. Agglomerative hierarchical clustering analyses were performed in the R statistical programming environment (version 2.13). The average linkage method was used in all cases. Because different data types were used in the various analyses, the metric used for clustering differed between the analyses. The Manhattan distance metric was used for the copy number data because the data were encoded as {-1, 0, 1}. The magnitudes of the CNAs were not considered, owing to a multitude of confounding factors, including tumour heterogeneity and ploidy. The Kendall rank correlation was used for the SNV frequency data because the data distributions were not normal. The Pearson correlation was used for the methylation data, which were normally distributed.

Identification of CpG hypermethylation events. Human genomic DNA was isolated from matching primary and metastatic medulloblastomas obtained from Johns Hopkins University, the Virginia Commonwealth University and New York University. An EZ DNA Methylation Kit (Zymogen Research) was used to bisulphite convert 500ng each sample. The recovered DNA was profiled on HumanMethylation27 BeadChips (Illumina) at The Centre for Applied Genomics (TCAG). Subsequently 27,578 CpG dinucleotides spanning 14,495 genes were analysed. The probe signal intensity was corrected by using BeadStudio 3.2.0 software (Illumina). The background normalization and differential methylation analyses were performed against fetal cerebella using the custom error model (Illumina). Cancer-specific DNA hypermethylation events were defined as those with a 30% increase in methylation in at least one medulloblastoma sample relative to an average methylation level (less than 50%) in normal fetal and adult cerebellum samples. Unsupervised clustering using Euclidian hierarchical clustering metrics was then performed on 2,503 data points that were filtered for cancer-specific hypermethylation events. The CpG methylation data are available from the Gene Expression Omnibus under accession number GSE34356.

Bisulphite sequencing of CpG promoter methylation. Representative examples of primary-tumour- and metastasis-specific methylation events were identified from normalized Illumina Hg27 data. Bisulphite PCR (BSP) primers were designed using the EpiDesigner tool (SEQUENOM) (<http://www.epidesigner.com/>) to encompass a genomic region flanking the Illumina Hg27 gene-specific probe. DNA (500ng) from the primary tumour and the corresponding metastases was bisulphite converted using an EZ DNA Methylation Kit. Following PCR optimization, 10ng bisulphite-converted DNA was used to amplify the genomic regions of interest. Amplicons were subcloned into the pCR2.1-TOPO vector (Invitrogen), and plasmid DNA from 10–12 colonies was extracted using a PureLink Quick Plasmid Miniprep Kit (Invitrogen). Sequencing was performed at TCAG using the M13 reverse primer, 5'-CAGGAAACAGCTATGAC-3'. The following primers were also used: *MLH1* Fw, 5'-TTGTTGGAATGTTATTTAT TATTTAGGA; *MLH1* Rv, 5'-CATAAATATCCACCAAAAAACCAAAA-3'; *MRPS21* Fw, 5'-TTTTTGGTTTGTGTTGATTGTTT-3'; *MRPS21* Rv, 5'-CAA ATCTCAAAAATCTATCCTTTCC-3'; *RBPI* Fw, 5'-GTAGGGGAGGTATAG GTAGGTTGTG-3'; *RBPI* Rv, 5'-CTTAATCAAACCCCTAAACAAAAA-3'; *WNK2* Fw, 5'-GTGTTTGGTTTATAGAGATGGA-3'; and *WNK2* Rv, 5'-AC TCTCTTAATCCCACTCTAC-3'. Details about the primer design and PCR protocols are available on request.

Alignment and variant calling for whole-exome sequencing. Standard manufacturers' protocols were used to perform target capture with a TruSeq Exome Enrichment Kit (Illumina) and sequencing of 100-bp paired-end reads on a HiSeq sequencing system (Illumina). Approximately 10 gigabases of sequence was generated for each subject such that >90% of the coding bases of the exome defined by the Consensus CDS (CCDS) project were covered by at least ten reads. Adaptor sequences and quality trimmed reads were removed by using the FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/), and then a custom script was used to ensure that only read pairs with both mates present were subsequently used. Reads were aligned to Hg19 with BWA1, and duplicate reads were marked using Picard (<http://picard.sourceforge.net/>) and excluded from downstream analyses. SNVs and short insertions and deletions (indels) were called using SAMtools (<http://samtools.sourceforge.net/>) Pileup and varFilter2 with the base alignment quality (BAQ) adjustment disabled and were quality filtered to require at least 20% of reads supporting the variant call. Variants were annotated using both ANNOVAR3 and custom scripts to identify whether they affected protein coding sequence and whether they had previously been seen in dbSNP131, the 1,000 Genomes pilot release (November 2010) or in approximately 160 exomes that had previously been sequenced at our centre.

SNV analysis of whole-exome sequencing data. For clustering analysis, an SNV frequency matrix was constructed by calculating frequencies from the read counts of the reference and the alternative nucleotide. The matrix was not standardized (that is, converted to z scores) before clustering, because the absolute SNV frequencies were of interest.

For Venn analysis, the samples were grouped into primary–metastasis sets, and the filtered SNVs were used to identify SNVs that are enriched in one sample compared with all other samples of the same set, as determined by the hypergeometric test (P value threshold = 0.05). For sets consisting of three or more samples (A, B and C), an SNV was considered to be enriched in samples A and B if the SNV was enriched in A compared with C alone and also enriched in B compared with C alone. SNVs that were not enriched in any sample or subset of samples were considered to be common SNVs. Many of these common SNVs probably represented germline SNVs specific to the patient.

Analysis of CpG promoter methylation data. The similarities between the patient-matched metastatic and primary tumour samples and among patient-matched metastatic tumour samples were determined by using Pearson correlation analysis. As Pearson's r values are not normally distributed, they were standardized by Fisher's z transformation. Subsequently, the correlations between the metastatic samples and the matched primary tumour samples were compared with the correlations among the patient-matched metastatic samples, using the paired heteroscedastic Student's t -test.

Clustering analysis was performed as described above. The methylation matrix was not standardized before clustering, as doing so would entail discarding crucial

information on the differences in the overall methylation profiles among samples or the average methylation among CpG promoters.

The stability of the CpG hypermethylation profile clusters was assessed using three methods. First, the clustering analysis was run for different numbers of CpG hypermethylation sites that varied most widely among samples. The partitions generated by each clustering run were compared with the reference partitions generated by the original clustering based on the 1,000 most variable hypermethylated CpG islands using the Jaccard similarity index. The same analysis was applied to a set of 100 background hypermethylation data matrices in which the sites are permuted independently in each sample. Second, the clustering analysis was performed for random subsamples of 1,000 sites, for 1,000 repeat runs. In each run, the resultant cluster was compared with the original cluster using the Jaccard index. Analysis on the original data matrix was compared with a set of 100 background matrices, permuted as described above. Third, the cluster stability was further assessed by bootstrap resampling of the samples using the pvclust R package (version 1.2).

28. Gnekow, A. K. Recommendations of the brain tumor subcommittee for the reporting of trials. *Med. Pediatr. Oncol.* **24**, 104–108 (1995).

IDH mutation impairs histone demethylation and results in a block to cell differentiation

Chao Lu^{1,2}, Patrick S. Ward^{1,2}, Gurpreet S. Kapoor³, Dan Rohle^{4,5}, Sevin Turcan⁴, Omar Abdel-Wahab^{4,6}, Christopher R. Edwards⁷, Raya Khanin⁸, Maria E. Figueroa⁹, Ari Melnick⁹, Kathryn E. Wellen², Donald M. O'Rourke^{3,10}, Shelley L. Berger⁷, Timothy A. Chan⁴, Ross L. Levine^{4,6}, Ingo K. Mellinghoff^{4,5,11} & Craig B. Thompson¹

Recurrent mutations in isocitrate dehydrogenase 1 (IDH1) and IDH2 have been identified in gliomas, acute myeloid leukaemias (AML) and chondrosarcomas, and share a novel enzymatic property of producing 2-hydroxyglutarate (2HG) from α -ketoglutarate^{1–6}. Here we report that 2HG-producing IDH mutants can prevent the histone demethylation that is required for lineage-specific progenitor cells to differentiate into terminally differentiated cells. In tumour samples from glioma patients, IDH mutations were associated with a distinct gene expression profile enriched for genes expressed in neural progenitor cells, and this was associated with increased histone methylation. To test whether the ability of IDH mutants to promote histone methylation contributes to a block in cell differentiation in non-transformed cells, we tested the effect of neomorphic IDH mutants on adipocyte differentiation *in vitro*. Introduction of either mutant IDH or cell-permeable 2HG was associated with repression of the inducible expression of lineage-specific differentiation genes and a block to differentiation. This correlated with a significant increase in repressive histone methylation marks without observable changes in promoter DNA methylation. Gliomas were found to have elevated levels of similar histone repressive marks. Stable transfection of a 2HG-producing mutant IDH into immortalized astrocytes resulted in progressive accumulation of histone methylation. Of the marks examined, increased H3K9 methylation reproducibly preceded a rise in DNA methylation as cells were passaged in culture. Furthermore, we found that the 2HG-inhibitable H3K9 demethylase KDM4C was induced during adipocyte differentiation, and that RNA-interference suppression of KDM4C was sufficient to block differentiation. Together these data demonstrate that 2HG can inhibit histone demethylation and that inhibition of histone demethylation can be sufficient to block the differentiation of non-transformed cells.

The fact that IDH mutations were identified in multiple cancers with disparate tissues of origin suggests that 2HG-producing mutant enzymes probably affect some fundamental cellular processes that facilitate tumour progression. To study the effects of IDH mutations, we collected and performed gene expression microarray analysis on tumour specimens from patients with grade II–III oligodendroglioma. Sequencing results revealed a high frequency of IDH mutations in oligodendroglioma (33 of the samples had the R132 IDH1 mutation, 2 had the R172 IDH2 mutation and 6 were wild type for IDH1/2). Supervised analysis found a statistically enriched gene signature in IDH-mutant samples (q value <10%, fold change >2; Fig. 1a and Supplementary Table 1) that was independent of tumour grade and recurrence status and survived multiple testing corrections. Gene-ontology analysis identified the regulation of astrocyte and glial differentiation

as the top two functional categories enriched in differentially expressed genes (Supplementary Table 2). We previously reported that IDH mutation may promote leukaemogenesis by expanding the haematopoietic progenitor cell population and impairing haematopoietic differentiation⁷, and that such a phenotype could be attributed at least in part to mutant IDH-induced inhibition of TET2, an α -ketoglutarate (α KG)-dependent enzyme potentially involved in DNA demethylation^{7,8}. Although DNA hypermethylation has been associated with IDH mutation in glioma samples⁹, no mutations in TET family members have been found in this disease. We explored the possibility that IDH mutation may affect additional α KG-dependent enzymes that contribute to the regulation of cell differentiation.

Histone lysine methylation is an integral part of the post-translational modifications of histone tails that are important for chromatin organization and regulation of gene transcription^{10–13}. *In vitro* 2HG can competitively inhibit a family of α KG-dependent Jumonji-C domain histone demethylases (JHDMs)^{14,15}. To determine whether IDH-associated changes in histone methylation could be observed in cells, we ectopically expressed wild-type or mutant IDH1 or IDH2 in 293T cells and found that mutant IDH1 or IDH2 led to a marked increase in histone methylation compared to the wild-type enzymes. Transient transfection of wild-type IDH2 can also lead to increased 2HG production⁷. In all of the samples, the magnitude of increase in methylation correlated with the intracellular 2HG levels produced by IDH transfection (Fig. 1b and Supplementary Fig. 1). To test whether histone lysine methylation was dysregulated in gliomas with IDH mutation, immunohistochemistry analysis of patient oligodendroglioma samples was performed for several well-characterized histone marks. Compared to tumours with wild-type IDH, there was a statistically significant increase in the repressive trimethylation of H3K9 (H3K9me3) and an increasing trend in trimethylation of H3K27 (H3K27me3) in tumours with IDH1 mutation (Fig. 1c). No statistically significant difference was seen in trimethylation of H3K4 (H3K4me3), a mark associated with active transcription (data not shown). These data suggested that IDH mutations might preferentially affect the regulation of repressive histone methylation marks *in vivo*.

As IDH mutations were associated with glial tumours of the 'proneural' phenotype¹⁶, we sought to determine whether the persistence of histone repressive marks promoted by mutant IDH was sufficient to block the differentiation of non-transformed cells. Upon stimulation with a differentiation cocktail, immortalized murine 3T3-L1 cells undergo extensive chromatin remodelling, resulting in their maturation into adipocytes¹⁷. 3T3-L1 cells transduced with R172K mutant IDH2 produced 2HG whereas cells transduced with either wild-type IDH2 or vector alone did not (Fig. 2a). All three cell types were then

¹Cancer Biology and Genetics Program, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA. ²Department of Cancer Biology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. ³Department of Neurosurgery, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. ⁴Human Oncology and Pathogenesis Program, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA. ⁵Department of Pharmacology, Weill Cornell Medical College, New York, New York 10065, USA. ⁶Leukemia Service, Department of Medicine, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA. ⁷Department of Cell and Developmental Biology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. ⁸Bioinformatics Core, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA. ⁹Division of Hematology/Oncology, Weill Cornell Medical College, New York, New York 10065, USA. ¹⁰Department of Pathology & Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. ¹¹Department of Neurology, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA.

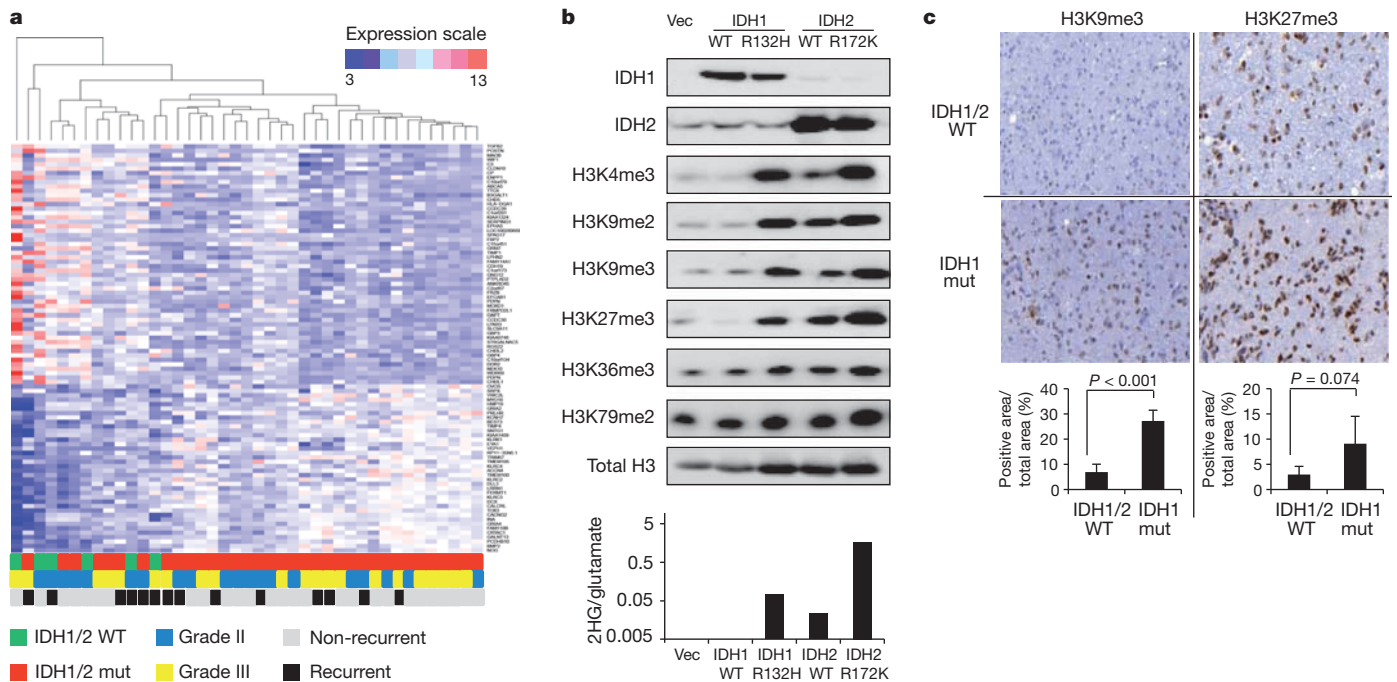


Figure 1 | IDH mutations are associated with dysregulation of glial differentiation and global histone methylation. **a**, Heatmap representation of a two-dimensional hierarchical clustering of genes identified as differentially expressed between IDH mutant (mut) patient oligodendroglioma samples and IDH wild-type (WT) samples. Each row represents a gene and each column represents a specimen. IDH mutational status, tumour grade and recurrence of each sample are listed. **b**, 293T cells transfected with empty vector (Vec), wild-type or R132H mutant IDH1, or wild-type or R172K mutant IDH2 for 3 days were lysed and assessed for expression levels of IDH1, IDH2 and histone lysine

induced to differentiate into adipocytes. After 7 days of differentiation induction, IDH mutant cells had visibly reduced lipid droplet accumulation compared to vector and IDH wild-type cells, as shown by Oil-Red-O staining (Fig. 2b). In separate experiments, stable transfection of R140Q mutant IDH2 also resulted in a block to adipocyte differentiation (data not shown). To determine whether 2HG was sufficient to mediate the effect of mutant IDH on cell differentiation, we synthesized cell-permeable 1-octyl-D-2-hydroxyglutarate (octyl-2HG; Supplementary Fig. 2). Treatment of 3T3-L1 cells with octyl-2HG led to a dose-dependent inhibition of lipid accumulation (Fig. 2c). Gene expression analysis showed that despite exposure to a well-standardized differentiation protocol¹⁸, IDH mutant cells or cells treated with octyl-2HG exhibited a profound defect in the expression of transcription factors essential for executing adipogenesis (*Cebpa* and *Pparg*) and an adipocytic lineage-specific gene (*Adipoq*) (Fig. 2d, e), suggesting that these cells failed to execute adipocyte differentiation.

Cells were harvested for a chromatin immunoprecipitation (ChIP) assay using antibodies against H3K9me3 and H3K27me3, before or after 4 days of differentiation induction. Quantitative polymerase chain reaction (PCR) with primers targeting promoters of *Cebpa* and *Adipoq* revealed that at day 4 there was a statistically significant increase in H3K9me3 and H3K27me3 at promoters of both genes in IDH mutant cells (Fig. 2f). These repressive marks also showed a modest but significant increase at gene promoters before differentiation induction. In contrast, quantitative assessment of DNA methylation at promoters of *Cebpa* and *Adipoq* by MassARRAY failed to reveal any significant difference between IDH wild-type and mutant cells (Supplementary Fig. 3). In addition to gene-specific changes, we detected a global increase in H3K9 methylation and a reciprocal decrease in H3 acetylation (Fig. 2g and Supplementary Fig. 4).

To determine whether IDH mutation was sufficient to induce enhanced repressive histone methylation in central nervous system

(CNS)-derived cells and whether it was associated with altered neural gene expression, we retrovirally transduced immortalized normal human astrocytes (NHAs) with either wild-type or R132H mutant IDH1. Compared to parental cells, late-passage cells expressing mutant IDH exhibited elevated levels of a variety of histone methylation marks (Fig. 3a), and this correlated with an enhanced expression of the neural marker nestin (Fig. 3b). IDH mutations have been associated with CpG-island hypermethylation⁹ and consistent with this we observed that total CpG methylation was increased in IDH mutant cells (Supplementary Fig. 5). Because histone repressive marks can promote DNA methylation and vice versa¹³, we studied the temporal relationship of histone and DNA methylation in IDH-expressing astrocytes (Fig. 3c–e and Supplementary Fig. 5). The first observable change of the histone marks we examined was H3K9me3. H3K9me3 levels were significantly elevated by passage 12 after cells were infected with mutant IDH. Changes in other histone methylation marks were either delayed and of lower magnitude (H3K27me3 and H3K79me2) or were not observed (H3K4me3). Increases in DNA methylation were never observed before passage 17 and the difference in DNA methylation reached statistical significance only at passage 22.

To test whether the IDH1 R132H mutation could interfere with neural differentiation in the absence of prolonged adaptation in culture, primary neurosphere cultures established from the brains of *p16/p19*^{-/-} mice were infected with a retroviral construct containing IDH1 R132H mutant, wild-type IDH1 or the vector alone (Supplementary Fig. 6). After infection the cells were re-plated under conditions that promote astrocyte differentiation and induced to differentiate further by treatment with retinoic acid without further passaging. IDH mutant cells failed to induce expression of the astrocytic marker GFAP and exhibited expression of the neural marker β 3-tubulin (Fig. 3f). When the differentiation conditions were supplemented with retinoic acid, enhanced expression of the astrocytic marker GFAP was observed in

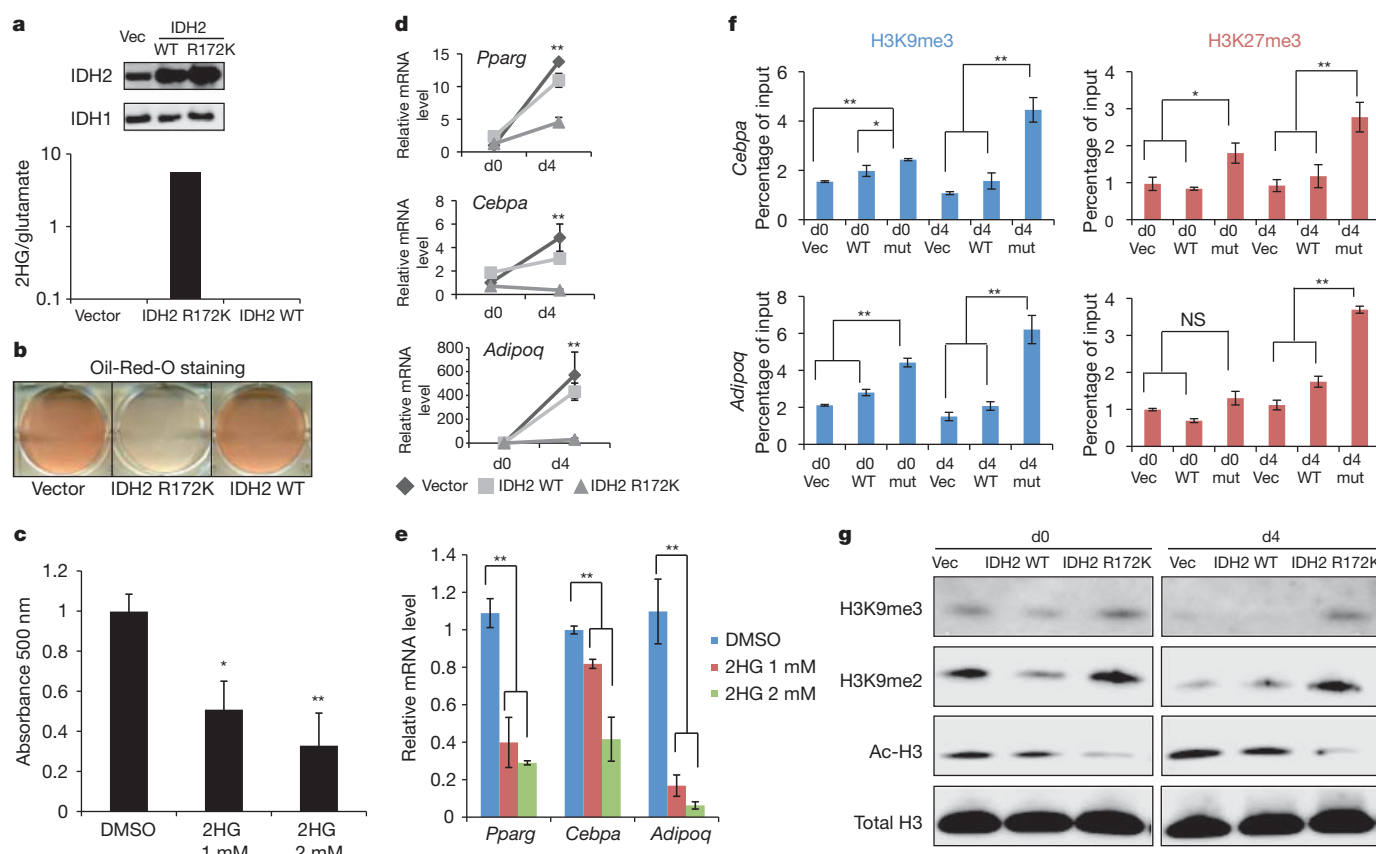


Figure 2 | Differentiation arrest induced by mutant IDH or 2HG is associated with increased global and promoter-specific H3K9 and H3K27 methylation. **a**, 3T3-L1 cells stably expressing empty vector, wild-type, or R172K mutant IDH2 were lysed and assessed for expression levels of IDH2 or IDH1 by western blotting. Cells were also extracted for intracellular metabolites, which were then MTBSTFA-derivatized (see Methods) and analysed by gas chromatography–mass spectrometry (GC–MS). The quantification of 2HG signal intensity relative to the intrasample glutamate signal is shown for a representative experiment. **b**, Cells were induced to differentiate into mature adipocytes for 7 days. The accumulation of lipid droplets was assessed by Oil-Red-O staining. Wells from a representative experiment from a total of four independent experiments are shown. **c**, 3T3-L1 cells were induced to differentiate for 7 days in the absence or presence of 1 mM or 2 mM octyl-2HG. Oil-Red-O staining was performed and quantified by measuring absorbance at 500 nm. DMSO, dimethylsulphoxide. **d**, Vector, wild-type or R172K mutant IDH2 transduced 3T3-L1 cells were induced to differentiate for 4 days. At days 0 and 4 (d0 and d4), RNA was extracted. Relative

IDH wild-type and vector cells but GFAP expression remained repressed in IDH mutant cells.

The enhancement of H3K9 methylation in mutant IDH-expressing cells from multiple tissues of origin led us to investigate whether this H3K9 methylation might be sufficient to block the ability of non-transformed cells to execute differentiation. Support for this hypothesis came with the discovery that KDM4C (also known as JMJD2C), an H3K9-specific JHDM, was induced in 3T3-L1 cells during differentiation (Fig. 4a). An *in vitro* histone demethylase assay with recombinant human GST-tagged KDM4C confirmed that KDM4C effectively removed H3K9me2 and H3K9me3 in the presence of α KG. Importantly, the demethylation reaction was inhibited by 2HG in a dose-dependent manner (Fig. 4b and Supplementary Fig. 7). Given the similarities between 2HG and α KG, the inhibition of KDM4C by 2HG would be predicted to be competitive. Consistently, increasing the concentration of α KG in the reaction mixture reversed the inhibition of H3K9 demethylation by 2HG (Fig. 4c).

Finally, to test the possibility that H3K9 demethylation is a required component of adipocyte differentiation, we examined whether blocking

expression of adipocyte-specific gene and transcription factors was assessed by quantitative PCR with reverse transcription (RT–qPCR). **e**, 3T3-L1 cells were induced to differentiate for 4 days in the absence or presence of 1 mM or 2 mM octyl-2HG. RNA was extracted. Relative expression of adipocyte-specific gene and transcription factors was assessed by RT–qPCR. **f**, Vector, wild-type or R172K mutant IDH2 transduced 3T3-L1 cells were induced to differentiate. At days 0 and 4 (d0 and d4), ChIP analysis was performed using antibodies against H3K9me3 and H3K27me3. Immunoprecipitated *Cebpa* and *Adipoq* promoter sequences were analysed by qPCR and shown as percentage of input. **g**, Vector, wild-type or R172K mutant IDH2 transduced 3T3-L1 cells were induced to differentiate for 4 days. At days 0 and 4 (d0 and d4), histones were acid-extracted and levels of H3K9me3, H3K9me2 and acetyl-H3 were assessed by western blotting with specific antibodies. Total H3 was used as loading control. Quantification of band intensities is shown in Supplementary Fig. 4. In **f**, error bars indicate s.d. from triplicate wells and a representative experiment from a total of two is shown. For all other experiments, error bars indicate s.d. from three independent experiments. * $P < 0.05$; ** $P < 0.01$; NS, not significant.

the induction of KDM4C was sufficient to impair the differentiation of 3T3-L1 cells. Treatment with three independent short interfering RNAs (siRNAs) against KDM4C reduced its expression and enhanced H3K9me3 in 3T3-L1 cells (Fig. 4d and Supplementary Fig. 8). After differentiation induction, cells treated with KDM4C siRNAs exhibited reduced ability to differentiate into adipocytes. Thus the inability to erase repressive H3K9 methylation can be sufficient to impair the differentiation of non-transformed cells.

Biochemical studies suggest that 2HG is a universal inhibitor of JHDM family members^{14,15}; therefore it was interesting to observe that H3K9 demethylation seemed to be more sensitive to mutant IDH-induced suppression than at least some other histone methylation marks. Future investigation of the sensitivity to 2HG inhibition among JHDM family members and/or cellular feedback mechanisms activated after defective histone demethylation will be needed. In addition to the data presented here, evidence is mounting for a direct role of histone methylation in stem cell maintenance, differentiation and tumorigenesis^{19–23} (see Supplementary Discussion). Our findings support a role for α KG-dependent demethylases in cell differentiation

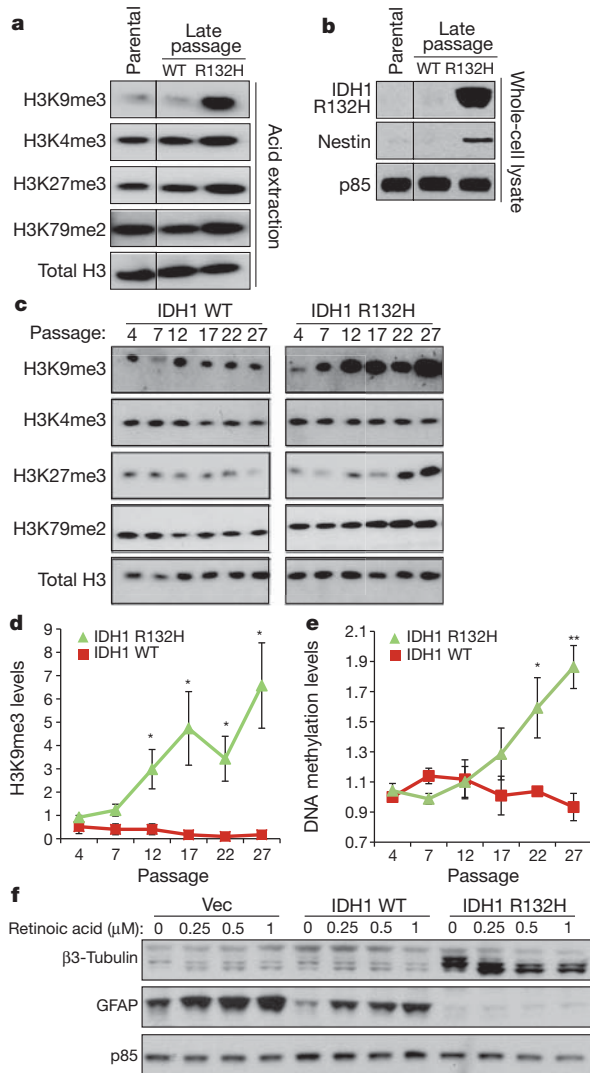


Figure 3 | IDH mutation induces histone methylation increase in CNS-derived cells and can alter cell lineage gene expression. **a**, Immortalized NHA cells were retrovirally transduced with constructs containing wild-type or R132H mutant IDH1. Histones were acid-extracted from parental cells or cells expressing wild-type or mutant IDH1 at late (>40) passages. Histone lysine methylation levels were assessed by western blotting with specific antibodies. Total H3 was used as loading control. Images presented are panels from different areas of the same gel. **b**, Parental, IDH1 wild-type and R132H mutant NHA cells at late passages were lysed and assessed for expression levels of nestin by western blotting. p85 was used as loading control. Images presented are panels from different areas of the same gel. **c**, NHA cells were retrovirally transduced with constructs containing wild-type or R132H mutant IDH1. Histones were acid-extracted at different time points as cells were passaged in culture. Histone lysine methylation levels were assessed by western blotting with specific antibodies. Total H3 was used as loading control. Images presented are panels from different areas of the same gel. **d**, Western blot band intensities of H3K9me3 in **c** and two additional independent experiments were quantified using Image J. Red squares indicate IDH1 wild-type cells. Green triangles indicate IDH1 R132H mutant cells. **e**, Total CpG methylation of IDH1 wild-type and R132H mutant NHA cells at various passages was measured by FACS using 5-methylcytosine-specific antibody and shown as normalized mean fluorescence intensity. FACS histograms from a representative experiment are shown in Supplementary Fig. 5. **f**, Neurosphere cultures established from the subventricular zone of brains of *p16/p19*^{-/-} mice were infected with a retroviral construct containing IDH1 R132H mutant, wild-type IDH1 or the vector alone and induced to differentiate. GFAP and β3-tubulin expression levels were assessed by western blotting. p85 was used as loading control. In **d** and **e**, error bars indicate standard error of the mean from three independent experiments. **P* < 0.05; ***P* < 0.01.

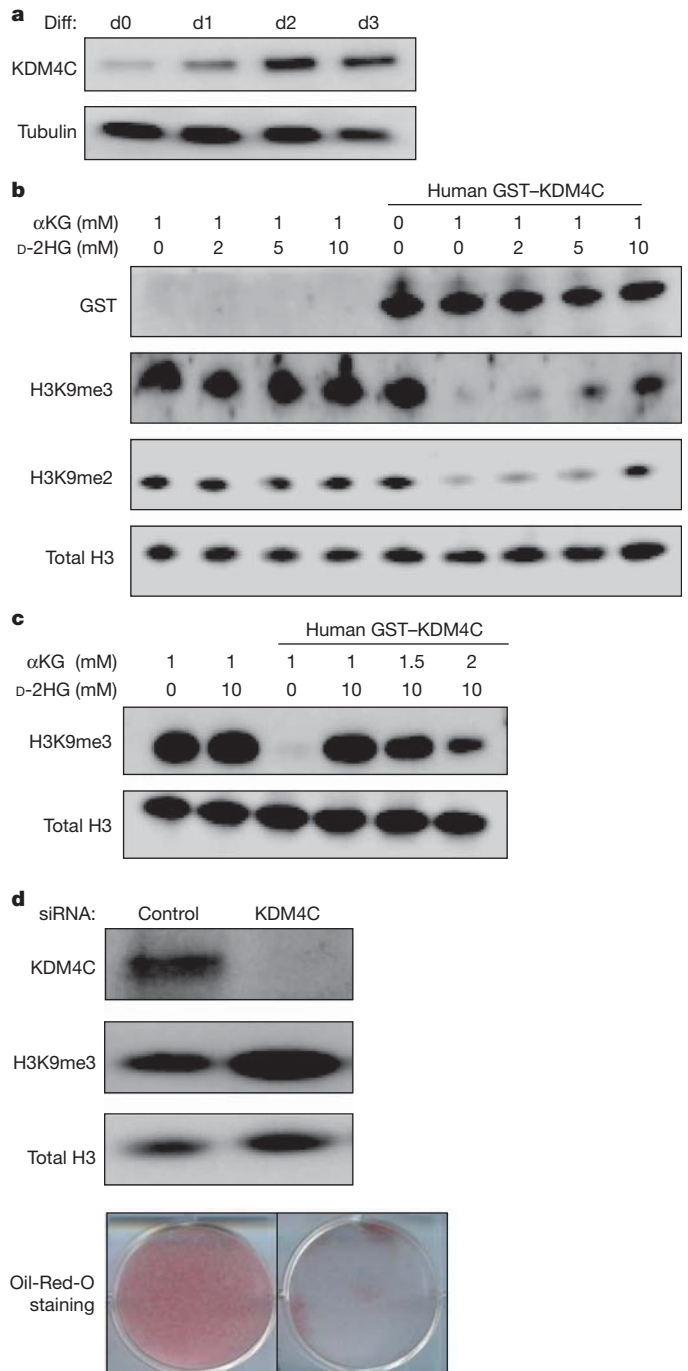


Figure 4 | 2HG-inhibitable H3K9 demethylase KDM4C is required for cell differentiation. **a**, 3T3-L1 cells were induced to differentiate (Diff) for 3 days. Before (d0) and each day after differentiation induction, cells were lysed and KDM4C protein levels were assessed by western blotting with specific antibody. Tubulin was used as loading control. **b**, Bulk histones were incubated with purified GST-tagged human KDM4C in the reaction mix with 1 mM αKG and increasing concentrations of D-2HG. Levels of GST tag, H3K9me3 and H3K9me2 were assessed by western blotting with specific antibodies. Total H3 was used as loading control. **c**, Bulk histones were incubated with purified GST-tagged human KDM4C in the reaction mix. 10 mM D-2HG was added to inhibit the demethylation reaction in the presence of increasing concentrations of αKG. Levels of H3K9me3 were assessed by western blotting with specific antibody. Total H3 was used as loading control. **d**, 3T3-L1 cells were transfected with control siRNA or siRNA specific for KDM4C. After 3 days, cells were lysed and assessed for expression levels of KDM4C and H3K9me3 by western blotting with specific antibodies. Total H3 was used as loading control. **e**, 3T3-L1 cells were induced to differentiate for 7 days. The accumulation of lipid droplets was assessed by Oil-Red-O staining. Wells from a representative experiment from a total of three independent experiments are shown.

that can be impaired through the cellular accumulation of 2HG produced by IDH mutation.

METHODS SUMMARY

Details about histone extraction, GC-MS and ChIP assay can be found in Methods. In brief, 3T3-L1 cell differentiation, Oil-Red-O staining, and *in vitro* histone demethylase assay were performed as previously described^{24,25}.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 3 June 2011; accepted 16 January 2012.

Published online 15 February 2012.

- Parsons, D. W. *et al.* An integrated genomic analysis of human glioblastoma multiforme. *Science* **321**, 1807–1812 (2008).
- Yan, H. *et al.* IDH1 and IDH2 mutations in gliomas. *N. Engl. J. Med.* **360**, 765–773 (2009).
- Mardis, E. R. *et al.* Recurring mutations found by sequencing an acute myeloid leukemia genome. *N. Engl. J. Med.* **361**, 1058–1066 (2009).
- Amary, M. F. *et al.* IDH1 and IDH2 mutations are frequent events in central chondrosarcoma and central and periosteal chondromas but not in other mesenchymal tumours. *J. Pathol.* **224**, 334–343 (2011).
- Ward, P. S. *et al.* The common feature of leukemia-associated IDH1 and IDH2 mutations is a neomorphic enzyme activity converting α -ketoglutarate to 2-hydroxyglutarate. *Cancer Cell* **17**, 225–234 (2010).
- Dang, L. *et al.* Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature* **462**, 739–744 (2009).
- Figuerola, M. E. *et al.* Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. *Cancer Cell* **18**, 553–567 (2010).
- Guo, J. U., Su, Y., Zhong, C., Ming, G. & Song, H. Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell* **145**, 423–434 (2011).
- Noushmehr, H. *et al.* Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma. *Cancer Cell* **17**, 510–522 (2010).
- Taverna, S. D., Li, H., Ruthenburg, A. J., Allis, C. D. & Patel, D. J. How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nature Struct. Mol. Biol.* **14**, 1025–1040 (2007).
- Berger, S. L. The complex language of chromatin regulation during transcription. *Nature* **447**, 407–412 (2007).
- Bernstein, B. E. *et al.* A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**, 315–326 (2006).
- Estève, P. O. *et al.* Direct interaction between DNMT1 and G9a coordinates DNA and histone methylation during replication. *Genes Dev.* **20**, 3089–3103 (2006).
- Xu, W. *et al.* Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of α -ketoglutarate-dependent dioxygenases. *Cancer Cell* **19**, 17–30 (2011).
- Chowdhury, R. *et al.* The oncometabolite 2-hydroxyglutarate inhibits histone lysine demethylases. *EMBO Rep.* **12**, 463–469 (2011).
- Verhaak, R. G. *et al.* Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in *PDGFRA*, *IDH1*, *EGFR*, and *NF1*. *Cancer Cell* **17**, 98–110 (2010).
- MacDougald, O. A. & Lane, M. D. Transcriptional regulation of gene expression during adipocyte differentiation. *Annu. Rev. Biochem.* **64**, 345–373 (1995).
- Rubin, C. S., Hirsch, A., Fung, C. & Rosen, O. M. Development of hormone receptors and hormonal responsiveness *in vitro*. Insulin receptors and insulin sensitivity in the preadipocyte and adipocyte forms of 3T3-L1 cells. *J. Biol. Chem.* **253**, 7570–7578 (1978).
- Krivtsov, A. V. & Armstrong, S. A. MLL translocations, histone modifications and leukaemia stem-cell development. *Nature Rev. Cancer* **7**, 823–833 (2007).
- Hu, Z. *et al.* A novel nuclear protein, 5qNCA (LOC51780) is a candidate for the myeloid leukemia tumor suppressor gene on chromosome 5 band q31. *Oncogene* **20**, 6946–6954 (2001).
- van Haften, G. *et al.* Somatic mutations of the histone H3K27 demethylase gene *UTX* in human cancer. *Nature Genet.* **41**, 521–523 (2009).
- Bilodeau, S., Kagey, M. H., Frampton, G. M., Rahl, P. B. & Young, R. A. SetDB1 contributes to repression of genes encoding developmental regulators and maintenance of ES cell state. *Genes Dev.* **23**, 2484–2489 (2009).
- Ceol, C. J. *et al.* The histone methyltransferase SETDB1 is recurrently amplified in melanoma and accelerates its onset. *Nature* **471**, 513–517 (2011).
- Wellen, K. E. *et al.* ATP-citrate lyase links cellular metabolism to histone acetylation. *Science* **324**, 1076–1080 (2009).
- Ingvarsdottir, K. *et al.* Histone H3K4 demethylation during activation and attenuation of *GAL1* transcription in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **27**, 7856–7864 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank members of the Thompson laboratory for technical help and critical reading of the manuscript. We thank T. A. Gocke and the genomic core of University of Pennsylvania for their assistance with the microarray study; the molecular cytology core facility of Memorial Sloan-Kettering Cancer Center (MSKCC) for technical help with the immunohistochemistry study; and O. Ouerfelli and M. K. Spassova at the organic synthesis core of MSKCC for the synthesis of octyl-2HG. This work was supported by grants from the National Cancer Institute and National Institutes of Health. R.L.L. is an Early Career Award recipient of the Howard Hughes Medical Institute and the Geoffrey Beene Junior Chair at MSKCC. D.M.O'R. is supported by the Betsy Cohen Fund of the Abramson Cancer Center at University of Pennsylvania. I.K.M. is supported by NCI-U54CA143798, the Doris Duke Charitable Foundation, and an Advanced Clinical Research Award in Glioma from the American Society of Clinical Oncology.

Author Contributions C.L., P.S.W. and C.B.T. designed the study. C.L., P.S.W., G.S.K., D.R. and M.E.F. performed research; S.T., D.R., T.A.C. and I.K.M. contributed research material; C.L., P.S.W., G.S.K., O.A.-W., C.R.E., R.K., M.E.F., A.M., K.E.W., D.M.O'R., S.L.B., R.L.L. and C.B.T. contributed to data analysis and interpretation; C.L., P.S.W. and C.B.T. wrote the manuscript.

Author Information Microarray data have been deposited with the ArrayExpress database under accession code E-MEXP-3239. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to C.B.T. (thompsonc@mskcc.org).

METHODS

Patient samples, microarray, gene-ontology analysis. Primary oligodendroglioma samples were obtained with approval from the institutional review board at the University of Pennsylvania and were de-identified for the study. For microarray analysis, tumour sample RNA was extracted with Trizol and purified with Qiagen RNeasy, and then assayed on an Affymetrix Human Gene 1.0ST array. Significance Analysis of Microarrays (<http://www-stat.stanford.edu/~tibs/SAM/sam.pdf>) was applied to find differentially expressed genes (q value <10% and fold change >2). Functional analysis of differentially expressed genes was done using the DAVID tool (<http://david.abcc.ncifcrf.gov/home.jsp>) using all human genes as a background set.

3T3-L1 cell differentiation, Oil-Red-O staining. 3T3-L1 cell differentiation and Oil-Red-O staining were carried out as described previously²⁴. In brief, confluent 3T3-L1 cells were stimulated with a cocktail containing 0.5 mM isobutylmethylxanthine, 1 μ M dexamethasone, 5 μ g/ml insulin and 5 μ M troglitazone (all from Sigma) to induce differentiation. Cells were maintained in medium with insulin after 2 days of differentiation until ready to be harvested. For Oil-Red-O staining, cells were washed in PBS and then fixed for 20 min at room temperature (25 °C) with 3% paraformaldehyde. Cells were then washed with de-ionized water and stained with Oil-Red-O solution. For quantification, Oil-Red-O staining was dissolved in isopropanol and absorbance was measured at 500 nm.

In vitro histone demethylase assay. The histone demethylase assay was carried out as described previously²⁵. In brief, 4 μ g bulk calf thymus histones (Sigma) were incubated with GST-tagged KDM4C (1.42 μ g; BPS Bioscience) in a reaction mix containing 50 mM Tris-HCl pH 8.0, protease inhibitors cocktail, 1 mM α KG, 100 μ M FeSO₄ and 2 mM ascorbic acid at 37 °C for 4 h, in the absence or presence of various concentrations of D-2HG or L-2HG (Sigma). Reaction mixtures were analysed by western blotting using specific antibodies.

Cell culture, transfection and transduction, generation of cell lines. 293T cells, NHA cells immortalized by E6/E7/hTERT (provided by R. Pieper²⁶) and 3T3-L1 cells were cultured in Dulbecco's modified Eagle's medium (DMEM; Invitrogen) with 10% fetal bovine serum (FBS; CellGro). For expression of wild-type and mutant IDH1/2 in 293T cells, transfection was performed with Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. For generation of IDH2 retrovirus and transduction of 3T3-L1 cells, supernatant from 293T cells transfected with pCL-Eco helper virus and plasmids was collected after 72 h, filtered and applied to cells overnight. For generating 3T3-L1 cell lines with stable expression of wild-type or mutant IDH2, cells were grown in 2.5 μ g ml⁻¹ puromycin for 7 days after retroviral transduction. Pooled populations of puromycin-resistant cells were obtained, and then continuously cultured in puromycin. For generation of IDH1 retrovirus and transduction of NHA cells, GP2-293 cells (Clontech, 631458) were calcium phosphate transfected with equal amounts of pVSV-G (Clontech, 631512) and plasmids. Virus was harvested at day 2 and day 3 after transfection and placed on logarithmically growing cells. After infection, cells were placed in 800 μ g ml⁻¹ G418 (Invitrogen) to generate stable cell lines. For siRNA knockdown of KDM4C, transfections were performed with Lipofectamine RNAiMAX (Invitrogen), using siRNAs targeting KDM4C (#1: sense, 5'-GCUUGAAUCUCCCAAGAUATT-3'; antisense, 5'-UAUCUU GGGAGAUUCAAGCTT-3'; #2: sense, 5'-CAAAGUAUCUUGGAUCAAAATT-3'; antisense, 5'-UUUGAUCCAAGAUACUUUGCC-3'; #3: sense, 5'-GAGGAGUU UCGGGAGUUCACAAAU-3'; antisense, 5'-AUUUGUUGAACUCCCGAA ACUCCUC-3') or a non-targeting control (Dharmacon, #D-001810-01-20) at a concentration of 40 nM.

Mutational analysis. For IDH mutation analysis, tumour genomic DNA was extracted and the regions surrounding IDH1 codon 132 and IDH2 codons 140 and 172 were amplified by PCR followed by sequencing. IDH1 analysis used forward primer 5'-ACCAAATGGCACCATACGA-3' and reverse primer 5'-TTCATACC TTGCTTAATGGGTGT-3' for amplification, and primer 5'-CGGTCTTCAGAG AAGCCATT-3' for sequencing¹. IDH2 analysis used forward primer 5'-CAG AGACAAGAGGATGGCTAGG-3' and reverse primer 5'-GTCTGCCTGTG TTGTTGCTTG-3' for amplification, and the same forward primer for sequencing²⁷. Out of the 42 tumours analysed, 41 had sufficient high quality genomic DNA for discerning IDH mutation status. The one sample unable to be classified as either IDH wild type or mutant was excluded from further analysis.

Plasmid construction. The cDNA clone of human IDH1 (BC012846.1) was purchased from the American Type Culture Collection in pCMV-Sport6, and human IDH2 (BC009244) was purchased from Invitrogen in pOTB7. Standard site-directed mutagenesis techniques were used to generate IDH1 R132H by introducing a G395A base-pair change in the IDH1 open reading frame (ORF). IDH2 R172K was made by introducing a G515A change in the IDH2 ORF, while IDH2 R140Q was made with a G419A alteration. Wild-type and mutant sequences were then subcloned into LPC vector. All sequences were confirmed by direct sequencing before expression in 293T cells and retrovirus generation. Retroviral constructs used for neurosphere infection were generated by excising wild-type

IDH1 and IDH1 R132H with NotI and PacI restriction enzymes from the previously made vectors and incorporating into the pQCXIH (Clontech, 631516) retroviral vector.

Histone extraction and western blotting. For histone acid extraction, cells were lysed in hypotonic lysis buffer (10 mM HEPES, 10 mM KCl, 1.5 mM MgCl₂, 0.5 mM DTT, protease inhibitors) for 1 h. H₂SO₄ was added to 0.2 N overnight at 4 °C with rotation. After spinning down and collecting supernatant, proteins were precipitated in 33% TCA, washed with acetone, and resuspended in de-ionized water. For whole-cell lysates, cells were lysed and sonicated in standard RIPA buffer (1% sodium deoxycholate, 0.1% SDS, 1% Triton X-100, 0.01 M Tris pH 8.0 and 0.14 M NaCl), and lysates were then centrifuged at 14,000g at 4 °C for 10 min. Supernatants were collected and measured for total protein concentration. For western blotting, lysates were separated by SDS-PAGE, transferred to nitrocellulose membrane, blocked in 5% non-fat milk in PBS containing 0.5% Tween-20, probed with primary antibodies and detected with horseradish-peroxidase-conjugated anti-rabbit or anti-mouse antibodies (GE Healthcare, NA934V and NA931V). Primary antibodies used were: anti-IDH1 (Proteintech, 12332-1-AP), anti-IDH2 (Abcam, ab55271), anti-GST tag (Millipore, 05-311), anti-H3K9me2 (Cell Signaling Tech, 9753), anti-H3K9me3 (Abcam, ab8898), anti-H3K36me3 (Abcam, ab9050), anti-H3K27me3 (Millipore, 17-622), anti-H3K4me3 (Millipore, 17-614), anti-H3K79me2 (Cell Signaling Tech, 9757), anti-KDM4C (Abcam, ab85454), anti-acetyl H3 (Upstate, 06-599), anti-H3 (Cell Signaling Tech, 4499), anti-tubulin (Sigma, T9026), anti-GFAP (Cell Signaling Tech, 3670), anti- β -tubulin (Cell Signaling Tech, 5666), anti-p85 (Millipore, 06-195), anti-nestin (Millipore, MAB5326), anti- β -actin (Sigma, A5316). Anti-IDH1 R132H mutant antibody was a gift from Agios Pharmaceuticals. Quantification of western blot band intensity was performed using Image J software according to the manufacturer's instructions.

Metabolite extraction, GC-MS. After gentle removal of culture medium, cells were rapidly quenched with ice-cold 80% methanol and incubated at -80 °C for 20 min. After sonication, extracts were then centrifuged at 14,000g for 20 min at 4 °C to remove precipitated protein and the aqueous metabolites in the supernatant layer were dried under nitrogen gas. For 293T cells, organic acids were further purified by redissolving the dried extract in de-ionized water, followed by elution from an AG-1 X8 100-200 anion exchange resin (Bio-Rad) in 3 N HCl after washing with five column volumes.

For GC-MS analysis, dried extracts were redissolved in a 1:1 mixture of acetonitrile and N-methyl-N-tert-butyltrimethylsilyltrifluoroacetamide (MTBSTFA; Regis) and heated for 75 min at 70 °C to derivatize metabolites. Samples were then injected into an Agilent 7890A GC with an HP-5MS capillary column, connected to an Agilent 5975 C mass selective detector operating in splitless mode using electron impact ionization with ionizing voltage of -70 eV and electron multiplier set to 1,060 V. GC temperature was started at 100 °C for 3 min, ramped up to 230 °C at 4 °C min⁻¹ and held for 4 min, then ramped up to 300 °C and held for 5 min. Mass range of 50-500 AMU was recorded at 2.71 scans per second. Identification of the 2HG metabolite peak was confirmed using standards obtained from Sigma. The 2HG and glutamate signal intensities were quantified by integration of peak areas.

Quantitative real-time PCR. RNA was isolated using Trizol (Invitrogen). After incubating with DNase, cDNA was synthesized using Superscript II reverse transcriptase (Invitrogen). Quantitative PCR was performed on a 7900HT Sequence Detection System (Applied Biosystems) using Taqman Gene Expression Assays (Applied Biosystems). Gene expression data was normalized to 18S rRNA.

ChIP. ChIP was performed with the Millipore Magna ChIP G kit (Millipore, 17-611). In brief, 2,000,000 cells were cross-linked with 1% formaldehyde for 10 min at room temperature. After washing with cold PBS, cells were centrifuged and lysed in 500 μ l SDS lysis buffer for 10 min on ice. Lysate was then sonicated using Bioruptor sonicator (Diagenode) to shear DNA to approximately 200-600 bp. Samples were spun down and 50 μ l of the supernatant was used for each immunoprecipitation overnight with magnetic beads after 10 \times dilution. Primary antibodies (3 μ g per ChIP) used were: anti-H3K9me3 (Abcam, ab8898) and anti-H3K27me3 (Millipore, 17-622). Normal rabbit IgG (Millipore, 12-370) was used as control and showed minimal enrichment. The next day, samples were washed in low-salt immune complex buffer, high-salt immune complex buffer, LiCl immune complex buffer and TE buffer. Histone complexes were eluted in elution buffer plus proteinase K for 2 h at 65 °C. DNA was recovered using columns. Quantitative PCR was performed on purified DNA samples. Primers used are: *Adipoq* forward, 5'-ATGGTGTAACACACAGCTTCA-3'; reverse, 5'-AGGGGTCAGGAGACCTCCCTTT-3'; *Cebpa* forward, 5'-CTGGAAGTGGGTGACTTAGAGG-3'; reverse, 5'-GAGTGGGGAGCATAGTGCTAG-3'. Data points (Ct) are converted to percentage of input.

Quantitative DNA methylation analysis. Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry using EpiTyper by MassARRAY

(Sequenom) was performed on bisulphite-converted DNA extracted from 3T3-L1 cells. MassARRAY primer design was done as previously described^{28,29}.

Immunohistochemistry. Immunohistochemistry detection was performed using Discovery XT processor (Ventana Medical Systems). The tissue sections were blocked for 30 min in 10% normal goat serum in 0.2% BSA/PBS, followed by incubation for 5 h with 0.1 $\mu\text{g ml}^{-1}$ of the rabbit polyclonal anti-H3K9me3 (Abcam, ab8898) or 1 $\mu\text{g ml}^{-1}$ rabbit polyclonal anti-H3K27me3 (Millipore, 07-449) antibodies and incubation for 60 min with biotinylated goat anti-rabbit IgG (Vector labs, PK6101) at 1:200 dilution. The detection was performed with the DAB-MAP kit (Ventana Medical Systems). The entire slides were scanned by Zeiss Mirax Scan (Carl Zeiss) using a $\times 20/0.8$ objective. The scanned image was exported into image analysis software, Metamorph (Molecular Devices). The colour threshold for DAB-positive nuclei was determined and set for all images. Areas above the threshold for the DAB signal and for haematoxylin-counterstained total nuclei were measured in an automated fashion. The ratio between the two parameters were calculated and analysed for statistical significance.

Synthesis of 1-octyl-D-2-hydroxyglutarate. Commercial *R*(-)-tetrahydro-5-oxofuran-2-carboxylic acid (140 mg, 1.076 mmol) was dissolved in H_2O (1 ml), cooled to 0 °C and treated with 1 N KOH (2.16 ml, 2.15 mmol). The resulting solution was stirred at this temperature for 5 min and at ambient temperature for 2 h. It was then concentrated to dryness under reduced pressure and dried. The residue was dissolved in trifluoroacetic anhydride (8 ml) at 0 °C, stirred for 30 min at 0 °C, for 2 h at room temperature, then the volatiles were evaporated under reduced pressure. The residue was dried and dissolved in anhydrous tetrahydrofuran (6 ml). Octanol (0.3 ml, 2.1 eq.) was added to the solution at 0 °C and the mixture was stirred for an overnight period at ambient temperature. Water was added to quench the reaction, and the mixture extracted with EtOAc. The combined extracts were dried over MgSO_4 , concentrated and purified by Flash chromatography (EtOAc:hexane 1:3 and 1:1) to give 1-octyl-D-2-hydroxyglutarate (110 mg, 39%).

Neurosphere isolation, culture and differentiation. Six days postpartum *Ink4a/Arf* null (*p16/p19*^{-/-}) mice were killed, with the isolated subventricular zones subjected to chemical (Pronase E, Calbiochem 7433-2) and mechanical dissociation to obtain a single-cell suspension in full neurobasal medium (Neurobasal medium, GIBCO 21103; B27 supplement without retinoic acid, GIBCO 12587-010;

Glutamax, GIBCO 35050; 20 ng ml^{-1} EGF, R&D Systems 236-EG; 20 ng ml^{-1} basic FGF, Millipore GF003). On the next day the cells were spun down and re-suspended in fresh medium, and once neurospheres had formed in culture, the spheres were collected and chemically dissociated (Accumax, Innovative Cell Technologies AM105) back into single cells in fresh medium.

One day after final infection, infected neurospheres and a non-infected control were placed in 400 $\mu\text{g ml}^{-1}$ Hygromycin B (InvivoGen, ant-hg-1). Once selection was complete, isogenic cell lines maintained in full neurobasal medium were chemically dissociated into single cells and plated at the same density in full neurobasal medium with increasing concentrations of retinoic acid (Sigma-Aldrich, R2625). Seventy-two hours later cells were harvested, and expression of proteins was analysed by western blotting.

Measurement of total CpG methylation. DNA methylation was assessed as previously described³⁰. In brief, 1×10^6 NHA cells were washed with PBS and fixed with 2% paraformaldehyde for 10 min at room temperature and permeabilized with 0.5% Triton X-100 for 10 min. Cells were then treated with 2 N HCl for 20 min at room temperature and subsequently neutralized with 100 mM Tris-HCl, pH 8.0. Cells were incubated with anti-5-methylcytosine antibody (Calbiochem, NA 81) at 1:100 dilution for 30 min at room temperature. After washing with PBS, cells were incubated with secondary antibody coupled with ALEXA FLUOR 488 (Invitrogen) for 30 min in the dark. Flow cytometry was done using Becton Dickinson Calibur flow cytometer and analysed using FlowJo software.

Statistical analysis. All statistical analysis was performed using Student's *t*-test (two-sample equal variance; two-tailed distribution).

26. Sonoda, Y. *et al.* Formation of intracranial tumors by genetically modified human astrocytes defines four pathways critical in the development of human anaplastic astrocytoma. *Cancer Res.* **61**, 4956–4960 (2001).
27. Balss, J. *et al.* Analysis of the *IDH1* codon 132 mutation in brain tumors. *Acta Neuropathol.* **116**, 597–602 (2008).
28. Figueroa, M. E. *et al.* Genome wide epigenetic analysis delineates a biologically distinct immature acute leukemia with myeloid/T-lymphoid features. *Blood* **113**, 2795–2804 (2009).
29. Figueroa, M. E. *et al.* DNA methylation signatures identify biologically distinct subtypes in acute myeloid leukemia. *Cancer Cell* **17**, 13–27 (2010).
30. Habib, M. *et al.* DNA global hypomethylation in EBV-transformed interphase nuclei. *Exp. Cell Res.* **249**, 46–53 (1999).

IDH1 mutation is sufficient to establish the glioma hypermethylator phenotype

Sevin Turcan^{1*}, Daniel Rohle^{1,2*}, Anuj Goenka^{1,3*}, Logan A. Walsh¹, Fang Fang¹, Emrullah Yilmaz¹, Carl Campos¹, Armida W. M. Fabius¹, Chao Lu^{4,5}, Patrick S. Ward^{4,5}, Craig B. Thompson⁴, Andrew Kaufman¹, Olga Guryanova¹, Ross Levine¹, Adriana Heguy¹, Agnes Viale⁶, Luc G. T. Morris^{1,7}, Jason T. Huse^{1,8}, Ingo K. Mellinghoff^{1,2,9,10} & Timothy A. Chan^{1,2,3,10}

Both genome-wide genetic and epigenetic alterations are fundamentally important for the development of cancers, but the interdependence of these aberrations is poorly understood. Glioblastomas and other cancers with the CpG island methylator phenotype (CIMP) constitute a subset of tumours with extensive epigenomic aberrations and a distinct biology^{1–3}. Glioma CIMP (G-CIMP) is a powerful determinant of tumour pathogenicity, but the molecular basis of G-CIMP remains unresolved. Here we show that mutation of a single gene, isocitrate dehydrogenase 1 (*IDH1*), establishes G-CIMP by remodelling the methylome. This remodelling results in reorganization of the methylome and transcriptome. Examination of the epigenome of a large set of intermediate-grade gliomas demonstrates a distinct G-CIMP phenotype that is highly dependent on the presence of *IDH* mutation. Introduction of mutant *IDH1* into primary human astrocytes alters specific histone marks, induces extensive DNA hypermethylation, and reshapes the methylome in a fashion that mirrors the changes observed in G-CIMP-positive lower-grade gliomas. Furthermore, the epigenomic alterations resulting from mutant *IDH1* activate key gene expression programs, characterize G-CIMP-positive proneural glioblastomas but not other glioblastomas, and are predictive of improved survival. Our findings demonstrate that *IDH* mutation is the molecular basis of CIMP in gliomas, provide a framework for understanding oncogenesis in these gliomas, and highlight the interplay between genomic and epigenomic changes in human cancers.

The isocitrate dehydrogenase genes *IDH1* and *IDH2* are mutated in >70% of lower-grade gliomas (grades II and III), in some glioblastomas^{4,5}, and in leukaemias and several other cancers^{6,7}. The most common *IDH1* mutations in glioma (>95%) result in an amino acid substitution at arginine 132 (R132), which resides in the enzyme's active site. Mutation of *IDH* imparts the ability to produce 2-hydroxyglutarate (2-HG), a potential oncometabolite^{8–10}. Alterations in the methylation landscape have been shown to have important roles during oncogenesis¹¹. CIMP has emerged as a distinct molecular subclass of tumours in a number of human malignancies, including glioblastoma^{1–3}. This phenotype is associated with extensive, coordinated hypermethylation at specific loci^{1,2,12,13}. In glioblastomas, G-CIMP is associated with the proneural subgroup of tumours and *IDH* mutation¹. Exactly how mutant *IDH* promotes tumorigenesis and causes G-CIMP—or CIMP in any type of human cancer—is unknown.

To determine whether *IDH1* mutation directly causes G-CIMP, we used immortalized primary human astrocytes¹⁴ and constructed isogenic cells expressing either mutant *IDH1* (R132H), wild-type *IDH1*, or neither. These astrocytes are well characterized^{14–17}. Introduction of

wild-type *IDH1* and the R132H *IDH1* mutant resulted in equal expression of protein (modest threefold increase) (Fig. 1a). Expression of mutant but not wild-type *IDH1* in human astrocytes resulted in the production of 2-HG (Fig. 1b). To determine whether mutant *IDH1* altered the methylation landscape, we analysed genomic DNA from these cells using the Illumina Infinium HumanMethylation450 platform. The platform provides genome-wide coverage and is both well validated and highly reproducible^{18,19}.

Previous data demonstrated that *de novo* DNA methylation in *in vitro* models occurs over extended periods, requiring time to 'lock in' epigenomic changes^{12,20}. We thus analysed the methylomes of astrocytes expressing mutant or wild-type *IDH1* over successive passages (up to 50). Analysis using self-organizing maps demonstrated that mutant *IDH1* progressively remodelled the glial methylome over time (Fig. 1c, d), an effect that was not seen in control astrocytes. Expression of mutant *IDH1* caused a marked increase in hypermethylation at a large number of genes, although there was a small group of hypomethylated genes as well (Fig. 1e and Supplementary Fig. 1a and Supplementary Table 1). Surprisingly, expression of wild-type *IDH1* also reshaped the methylome but in a manner that differed from effects due to expression of mutant *IDH1* (Fig. 1f). Expression of wild-type *IDH1* caused hypomethylation at specific loci, suggesting that both the production of 2-HG and the levels of α -ketoglutarate can affect the methylome. Unsupervised hierarchical clustering of the methylome data showed that the hypermethylated genes included both genes that underwent *de novo* methylation as well as genes that originally possessed low levels of methylation but subsequently acquired high levels of methylation (Fig. 1e). Control astrocytes did not undergo these methylome changes (Fig. 1c, d). Mutant *IDH1*-induced remodeling of the methylome was progressive and reproducible, and resulted in significant changes in gene expression (Fig. 1f and Supplementary Fig. 1a, Supplementary Tables 2 and 3).

We sought to define the methylation targets of mutant *IDH* in astrocytes. Of the 44,334 CpG sites that were differentially methylated in mutant *IDH*-expressing cells, 30,988 sites were hypermethylated (3,141 unique genes with promoter CpG island methylation changes; Supplementary Table 1). Transcriptional module mapping showed that the genes undergoing methylation changes were highly enriched for polycomb complex 2 (PRC2)-targeted loci (Supplementary Fig. 1b and Supplementary Table 4)^{12,21}. These observations demonstrate that mutant *IDH1* is sufficient to reshape the epigenome by altering the global methylation landscape.

Lower-grade gliomas (LGGs; World Health Organization grades II and III) and secondary glioblastomas are biologically distinct from primary or *de novo* glioblastomas²². Present knowledge of G-CIMP

¹Human Oncology and Pathogenesis Program, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA. ²Weill Cornell College of Medicine, New York, New York 10065, USA. ³Department of Radiation Oncology, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA. ⁴Cancer Biology and Genetics, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA. ⁵Department of Cancer Biology, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104, USA. ⁶Genomics Core, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA. ⁷Department of Surgery, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA. ⁸Department of Pathology, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA. ⁹Department of Neurology, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA. ¹⁰Brain Tumor Center, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA.

*These authors contributed equally to this work.

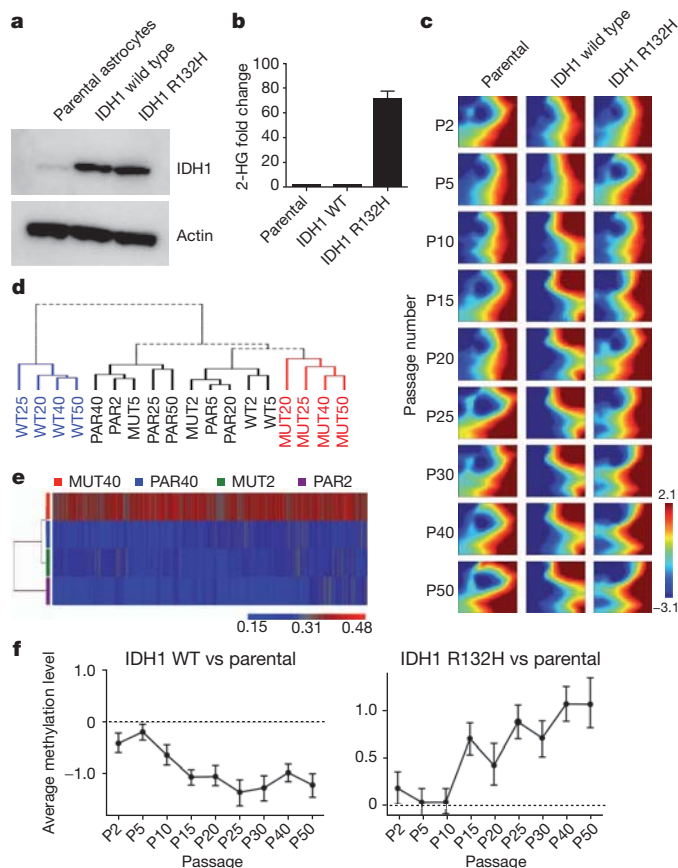


Figure 1 | Introduction of mutant IDH1 into human astrocytes remodels the methylome. **a**, Expression of wild-type and mutant IDH1 (R132H) in immortalized human astrocytes (passage 5). **b**, Overexpression of mutant IDH1 but not wild-type (WT) IDH1 in human astrocytes leads to production of 2-HG⁸. Error bars show 1 standard deviation (s.d.). ($n = 2$). **c**, Self-organizing map (SOM) analysis of methylome data for wild-type IDH1-expressing, mutant IDH1-expressing (R132H), and parental (control) cell lines shows changes in the methylome in mutant IDH1-expressing and wild-type IDH1-expressing astrocytes, compared to parental cells. Mosaic patterns are pseudo-coloured SOMs from different time points (P indicates passage number). Tile colours indicate methylation level of centroids. **d**, Hierarchical clustering showing divergence of the methylome of IDH1-expressing astrocytes from that of parental astrocytes. MUT, mutant; PAR, parental. **e**, Heatmap showing the 10,678 most significant differentially methylated probes (ANOVA) in IDH1 mutant astrocytes and parental astrocytes (passages 2 and 40). Colour scale indicates β values. **f**, Kinetics of differential methylation in mutant and wild-type-expressing astrocytes. Error bars indicate inter-quartile range ($n = 2$).

is based on the examination of primary glioblastomas in which IDH mutations are infrequent^{1,4,5}. To determine the impact of IDH mutation on the methylation landscape in primary LGGs, we generated a high-resolution, genome-wide set of LGG methylome data from patients with complete clinical follow-up using the same Infinium 450K platform as described earlier (72 WHO grade II and III gliomas; Fig. 2 and Supplementary Table 5). We first performed consensus clustering (Fig. 2a and Supplementary Fig. 2a) and unsupervised hierarchical clustering (Fig. 2b and Supplementary Fig. 2b) to identify LGG subgroups. We identified two robust DNA methylation clusters, one encompassing tumours with markedly high methylation levels (cluster 2) and another without the hypermethylated loci (cluster 1). Cluster 2 tumours demonstrated a characteristic DNA methylation profile with high-coordinate cancer-specific methylation at a subset of loci, concordant with the G-CIMP phenotype defined in glioblastomas (Supplementary Fig. 2b and Supplementary Table 6)¹. The composition of the G-CIMP group in these LGGs was confirmed by two independent clustering methods (*K*-means consensus and two-dimensional

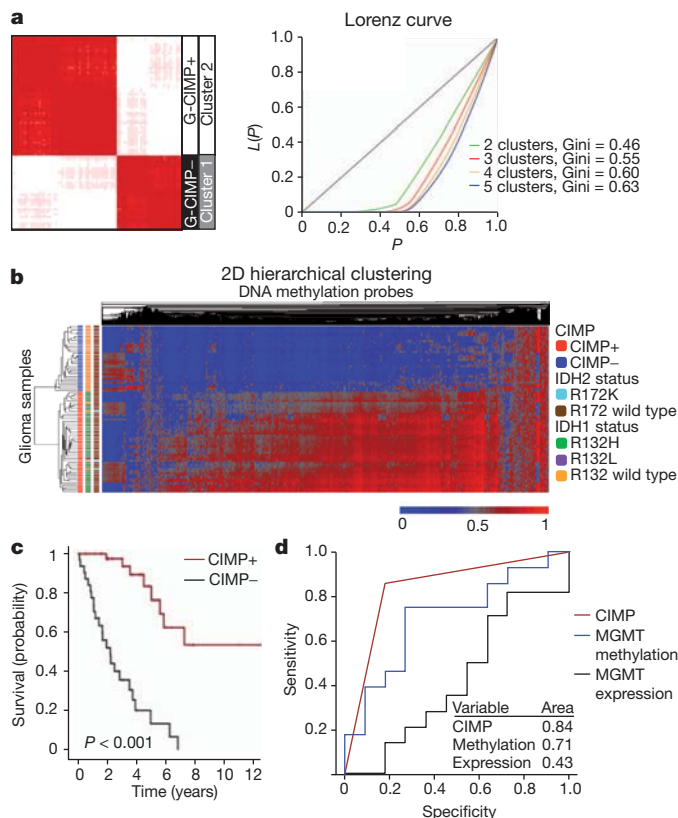


Figure 2 | Global epigenetic analysis of LGGs reveals dependence of G-CIMP on IDH mutation. **a**, Identification of G-CIMP by *K*-means consensus clustering of LGG samples. Unsupervised clustering was performed with the most variant probes (9,711 probes, top 2%). Tumours are listed in the same order along the x and y axes. G-CIMP status is indicated by the black and white bars. Consensus index values range from 0 to 1, with 0 being dissimilar (white) and 1 being similar (red). $K = 2$ is identified by the Lorenz curve. **b**, Two-dimensional (2D) hierarchical clustering of the same probes as in **a** identified the same two clusters. Each row represents a tumour and each column represents a probe. CIMP and IDH mutation status are indicated by the colour code. The level of DNA methylation (β value) for each probe is represented by colour scale (red, methylated; blue, non-methylated). Only cancer-specific events were used²⁷. **c**, Kaplan-Meier survival curve of Memorial Sloan-Kettering Cancer Center (MSKCC) patients ($n = 72$) with LGG (grade II and III). $P < 0.001$. **d**, Receiver operating characteristic (ROC) curve comparing the sensitivity and specificity of G-CIMP status compared with MGMT methylation or MGMT expression status, in LGGs. Areas under the curve are noted in the inset. G-CIMP, MGMT methylation and MGMT expression were determined as described in Methods.

hierarchical clustering) (Fig. 2a, b). Probes defining CIMP in LGGs included those in CpG islands and shores (Supplementary Fig. 2c, d) and were enriched for PRC2-target genes (Supplementary Table 7). Global expression profiles showed that G-CIMP+ tumours possessed markedly different transcriptional profiles than G-CIMP- tumours (Supplementary Tables 8 and 9). EpiTYPER (Sequenom) mass spectrometry was used to validate the methylation status of loci in both the astrocyte model and in the tumours (Supplementary Fig. 2e–g)²³.

To determine the mutational status of *IDH1* and *IDH2*, we sequenced the entire coding sequence of the two genes in all the samples above (Fig. 2b). Ninety-eight per cent (49/50) of the G-CIMP+ tumours possessed either an *IDH1* mutation or *IDH2* mutation. Notably, none of the G-CIMP- tumours possessed mutant IDH (Supplementary Fig. 2h). These genomic data show that G-CIMP is highly dependent on the presence of IDH mutation and, in LGGs that are CIMP-, IDH mutations do not occur (0%). Currently, the methylation status of O-6-methylguanine DNA methyltransferase (MGMT) is a widely used molecular biomarker for glioblastoma prognosis and response to

temozolomide²⁴. In LGGs, G-CIMP associated with markedly better clinical endpoints (Fig. 2c and Supplementary Figs 3–6, Supplementary Tables 10 and 11). Importantly, G-CIMP was significantly superior to MGMT methylation or MGMT messenger RNA expression as a predictor of survival (Fig. 2d).

We next sought to define the nature of the methylome differences between IDH mutant and wild-type tumours and characterize the effects of these differences on the LGG transcriptome. Figure 3a shows a principal component analysis (PCA) of methylome and expression data from our tumours. PCA shows that G-CIMP+ and G-CIMP– LGGs methylome subgroups correlate with marked transcriptome differences (Fig. 3a). Of the 140,016 sites that were differentially methylated between IDH mutant and wild-type tumours, 121,660 were hypermethylated (Supplementary Table 6). There were 2,611 unique genes with alterations in promoter CpG islands represented in this group. Consistent with the results in Fig. 2b, a volcano plot showing differentially methylated genes between G-CIMP+ and G-CIMP– tumours was highly asymmetric (Fig. 3b). A starburst plot showing the relationship between DNA methylation and expression is shown in Fig. 3b. Integration of the normalized gene expression and DNA methylation gene sets identified 429 genes with both significant hypermethylation and downregulation and 176 genes that were hypomethylated and upregulated in G-CIMP+ LGGs (Supplementary Table 12). Among these genes are those known to be involved in glioma initiation and outcome, including *CDKN2C* and *GAP43* (refs 25, 26).

As a critical experiment to prove causality between IDH1 mutation and G-CIMP, we performed an in-depth comparison of methylation marks and gene expression alterations between human astrocytes expressing mutant IDH1 and the LGGs with endogenous IDH1 mutation. We first focused on the comparison of methylation marks and found that both sets of methylome alterations targeted similar loci. Gene set enrichment analysis (GSEA) of the mutant IDH1-induced methylation changes in the isogenic astrocyte system (Fig. 1) and the G-CIMP genes demonstrated very significant enrichment and concordance (Fig. 3c and Supplementary Table 13 and Supplementary Fig. 7). Importantly, the genes that were methylated after mutant IDH1 expression correctly classified LGG tumours into CIMP+ or CIMP– groups with very high accuracy (Fig. 3d and Supplementary Table 14). To confirm the impact of these alterations on glioma pathobiology, we used the transcriptomic footprint of mutant IDH to generate an expression signature (mutant IDH repression signature) composed of the most significantly methylated and downregulated genes in both the isogenic astrocyte system and the G-CIMP gene set (17 genes; Supplementary Table 15). As expected, this signature classified an independent LGG cohort (Rembrandt) into two distinct subgroups (Fig. 3e and Supplementary Figs 8–10 and Supplementary Table 16). Together, our findings show that introduction of mutant IDH reprograms the epigenome and generates the foundations of G-CIMP.

IDH mutation is highly enriched in the CIMP+, proneural subgroup of glioblastomas. Using data from The Cancer Genome Atlas (TCGA), we applied the mutant IDH repression signature as a classifier to the transcriptomes of all four subgroups of glioblastomas²⁷. The signature segregated IDH mutant and wild-type proneural glioblastomas into two distinct subgroups associated with very different prognoses, but did not do so in other glioblastoma subgroups (Supplementary Fig. 11a, b). These data demonstrate that mutant IDH-induced epigenomic alterations have profound biological implications within the proneural class of glioblastomas that are specific for this subclass. Comparison of gene expression programs that occur in astrocytes expressing mutant IDH1 to those in LGG tumours that harbour the IDH mutation showed remarkable similarity (Fig. 4a and Supplementary Fig. 12). Moreover, introduction of mutant but not wild-type IDH1 into astrocytes resulted in the upregulation of nestin (and other genes associated with stem cell identity) at the time of DNA methylation increase and the adoption of a neurosphere/stem-like phenotype (Fig. 4b and Supplementary

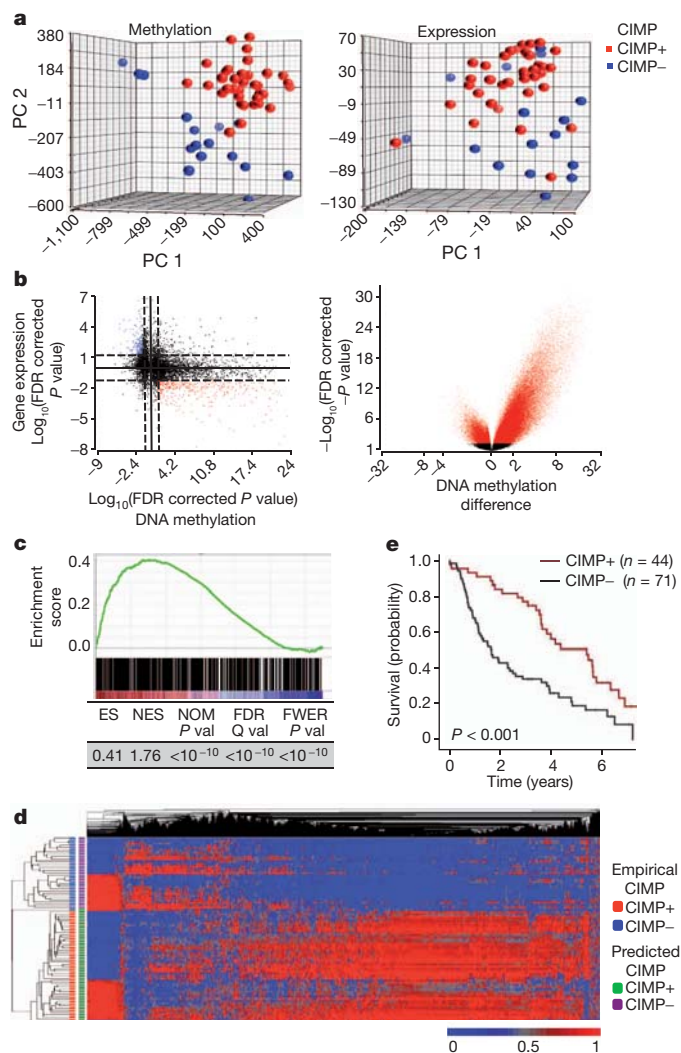


Figure 3 | IDH1 mutation directly generates the methylation patterns present in G-CIMP tumours. **a**, The methylomes and transcriptomes of LGGs are distinct. PCA plot of LGG tumours for all methylation probes (left) and expression probes (right) ($n = 52$). PC, principal component. **b**, Starburst plot (left) for comparison of DNA methylation and gene expression. The $\log_{10}(\text{FDR-corrected } P \text{ value})$ is plotted for β value for DNA methylation (x axis) and gene expression (y axis) for each gene. Black dotted line shows the FDR-adjusted P value of 0.05. Red points indicate downregulated and hypermethylated genes in G-CIMP+ LGGs versus G-CIMP– LGGs. Blue points show hypomethylated and upregulated genes. Volcano (right) plot of all CpG loci analysed for G-CIMP association. The β -value difference in DNA methylation between G-CIMP+ and G-CIMP– tumours is plotted along the x axis. The P value between G-CIMP+ and G-CIMP– tumours is plotted on the y axis ($-\log_{10}$ scale). Red indicates significantly different probes. **c**, Concordance between hypermethylated sites in mutant IDH1-expressing astrocytes and G-CIMP+ LGGs. GSEA shows significant enrichment between 730 hypermethylated unique CpG sites identified in IDH1 mutant astrocytes (ANOVA between passage 2 and 40) and those present in G-CIMP+ gliomas. GSEA correlation shown in colour scale. ES, enrichment score; FDR, false discovery rate; FEWR, familywise error rate; NES, normalized enrichment score; NOM, nominal P value. **d**, Differential methylation in IDH mutant astrocytes correctly classifies G-CIMP in the human LGGs. Two-dimensional-unsupervised hierarchical clustering of 81 human gliomas with top variant probes ($n = 10,000$) from mutant IDH1 astrocytes. Tumours are shown on the y axis, probes along the x axis. Methylation (β value) for each probe is represented with the colour scale. G-CIMP classification as determined by the astrocyte-derived data is denoted by the colour bars at the left. **e**, Kaplan–Meier survival curve of 115 patients with grade II or grade III gliomas in the Rembrandt Database grouped by CIMP status. P value calculated by log rank.

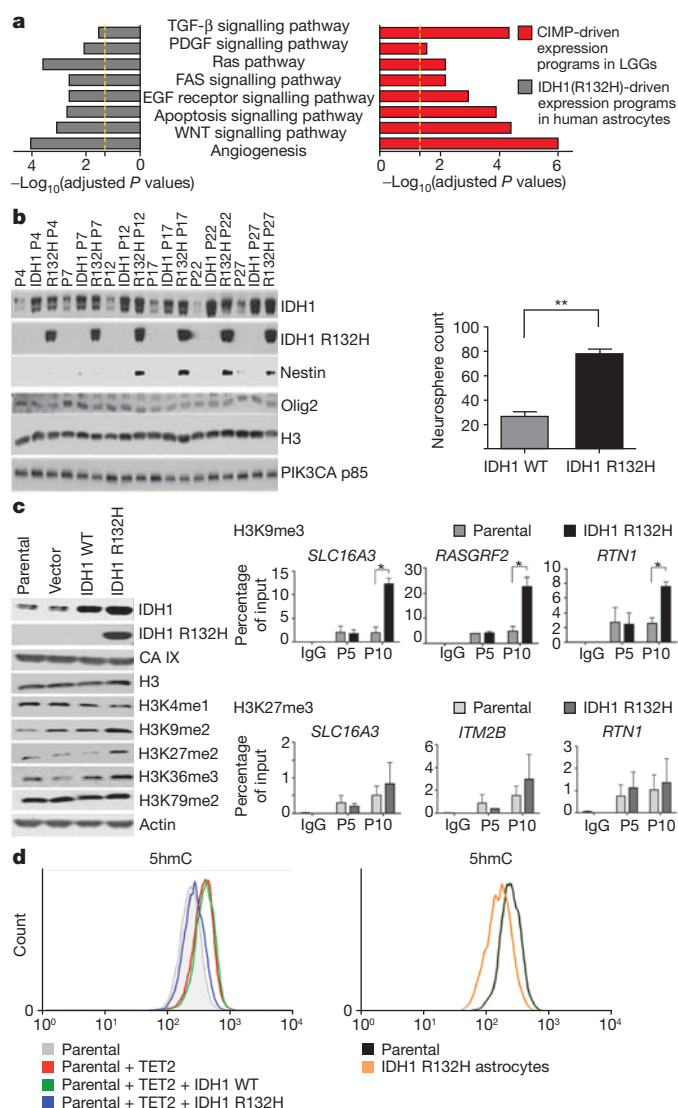


Figure 4 | Functional implications of IDH1-mutation-induced alterations in the glioma epigenome. **a**, Concordance of transcriptional programs regulated by mutant IDH1 in astrocytes and G-CIMP in LGGs. P value for significance is shown along the x axis. Yellow lines indicate threshold of significance ($P = 0.05$). **b**, Mutant IDH1 results in the expression of markers of self-renewal and stem cell identity. Left, mutant IDH1 results in expression of nestin. P indicates passage number. Right, expression of mutant IDH1 promotes the adoption of a neurosphere phenotype. Astrocytes (passage 15) that express IDH1 R132H or IDH1 wild type were used in the neurosphere assay. Error bars indicate 1 s.d. $^{**}P < 0.01$ (t -test). **c**, Alterations in histone marks in IDH1-mutant-expressing human astrocytes. Left, western blot results are shown using the indicated antibodies. Astrocytes are from passage 27. Right, ChIP of the indicated histone marks for representative hypermethylated genes. Error bars indicate 1 s.d. $^{*}P < 0.05$. **d**, Mutant IDH1 inhibits the production of 5hmC in human astrocytes. Left, mutant IDH inhibits TET2-dependent 5hmC production in astrocytes. Parental astrocytes were infected with lentivirus directing the expression of TET2 catalytic domain and green fluorescent protein (GFP) \pm mutant IDH1. FACS analyses are shown for 5hmC. Right, astrocytes expressing IDH R132H (passage 10) have less 5hmC than astrocytes that do not express the IDH mutant.

Fig. 13)²⁸. These data suggest that mutant IDH1 functions by interfering with differentiation state.

Our data show that IDH1 mutation is the mechanistic cause of G-CIMP. To gain further insight, we determined the effects of mutant IDH1 on histone alterations in our astrocyte system. Figure 4c (left) shows that expression of the IDH1 mutant increases levels of H3K9me2, H3K27me3 and H3K36me3, consistent with previous findings²⁹.

Chromatin immunoprecipitation (ChIP) experiments examining representative genes that undergo hypermethylation show H3K9 and H3K27 methylation are both enriched in cells expressing mutant IDH1 (Fig. 4c, right). As both of these marks can promote DNA methylation, alterations in histone marks may contribute to the accumulation of DNA methylation.

Next, we determined the effects the mutation had on TET2-dependent 5-hydroxymethylcytosine (5hmC) levels. We used a well-established assay^{9,29} and first confirmed that we were able to detect TET-dependent alterations in 5hmC (Supplementary Fig. 14). We found that expression of the IDH1 mutant in astrocytes resulted in a significant decrease in 5hmC (Fig. 4d, right). Expression of TET2 in the astrocytes produced 5hmC, which was inhibited by mutant but not wild-type IDH1 (Fig. 4d, left). Because TET-mediated production of 5hmC is a primary mode of DNA demethylation³⁰, inhibition of this activity in the IDH1-mutant-expressing astrocytes may be a mechanistic basis for accumulation of DNA methylation, ultimately leading to a CIMP pattern.

IDH mutation and the CIMP phenotype are two very common features in cancer, the underlying mechanisms for which are obscure. The fundamental questions regarding these features are (1) how the IDH mutation contributes to oncogenesis, and (2) what the root cause of CIMP is. Our data address these important questions by demonstrating that IDH mutation is the cause of CIMP and leads to the CIMP phenotype by stably reshaping the epigenome. This remodelling involves modulating patterns of methylation on a genome-wide scale, changing transcriptional programs and altering the differentiation state. Our observations suggest that the activity of IDH may form the basis of an 'epigenomic rheostat', linking alterations in cellular metabolism to the epigenetic state. In summary, these data provide a mechanistic framework for how IDH mutation leads to oncogenesis and the molecular basis of CIMP in gliomas. We believe our observations have critical implications for the understanding of gliomas and the development of novel therapies for this disease.

METHODS SUMMARY

Cell culture. Immortalized human astrocytes were a gift from R. O. Pieper (University of California, San Francisco) and were prepared as previously described¹⁴. Cells were cultured in Dulbecco's modified Eagle's medium (DMEM) plus 10% fetal bovine serum (FBS; Invitrogen). Expression of IDH was accomplished by cloning wild-type or mutant IDH1 (R132H) into the vector pLNCX2. These constructs were used to construct lentiviruses used for infection of target cells. Selection was performed using G418. All experiments were performed in duplicate.

Tumours. All tumours ($n = 81$) were obtained following surgical resection at the MSKCC as part of routine clinical care and snap frozen. Tumours were obtained in accordance with Institutional Review Board policies at the MSKCC. Each sample was examined histologically with haematoxylin-and-eosin-stained cryostat sections by a neuropathologist. Before analysis, tumours were sectioned and microdissected. Genomic DNA or RNA was extracted using the DNeasy kit (Qiagen) or RNeasy Lipid Tissue Mini kit (Qiagen) per the manufacturer's instructions.

Genomic analysis. Expression analysis of astrocytes and tumours was performed using the Affymetrix U133 2.0 microarray. Genome-wide methylation analysis was performed using the Illumina Infinium HumanMethylation450 bead array. Processing of the arrays was as per the manufacturer's protocol. Methylation data were extracted using GenomeStudio software (Illumina). Methylation values for each site are expressed as a β value, representing a continuous measurement from 0 (completely unmethylated) to 1 (completely methylated). This value is based on the following calculation: β value = (signal intensity of methylation-detection probe) / (signal intensity of methylation-detection probe + signal intensity of non-methylation detection probe).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 3 June 2011; accepted 17 January 2012.

Published online 15 February 2012.

1. Noshahr, H. *et al.* Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma. *Cancer Cell* **17**, 510–522 (2010).
2. Fang, F. *et al.* Breast cancer methylomes establish an epigenomic foundation for metastasis. *Sci. Transl. Med.* **3**, 75ra25 (2011).

3. Toyota, M. *et al.* CpG island methylator phenotype in colorectal cancer. *Proc. Natl Acad. Sci. USA* **96**, 8681–8686 (1999).
4. Yan, H. *et al.* IDH1 and IDH2 mutations in gliomas. *N. Engl. J. Med.* **360**, 765–773 (2009).
5. Parsons, D. W. *et al.* An integrated genomic analysis of human glioblastoma multiforme. *Science* **321**, 1807–1812 (2008).
6. Sjöblom, T. *et al.* The consensus coding sequences of human breast and colorectal cancers. *Science* **314**, 268–274 (2006).
7. Mardis, E. R. *et al.* Recurring mutations found by sequencing an acute myeloid leukemia genome. *N. Engl. J. Med.* **361**, 1058–1066 (2009).
8. Dang, L. *et al.* Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature* **462**, 739–744 (2009).
9. Figueroa, M. E. *et al.* Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. *Cancer Cell* **18**, 553–567 (2010).
10. Ward, P. S. *et al.* The common feature of leukemia-associated IDH1 and IDH2 mutations is a neomorphic enzyme activity converting α -ketoglutarate to 2-hydroxyglutarate. *Cancer Cell* **17**, 225–234 (2010).
11. Jones, P. A. & Baylin, S. B. The epigenomics of cancer. *Cell* **128**, 683–692 (2007).
12. Ohm, J. E. *et al.* A stem cell-like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing. *Nature Genet.* **39**, 237–242 (2007).
13. Cheng, Y. W. *et al.* CpG island methylator phenotype associates with low-degree chromosomal abnormalities in colorectal cancer. *Clin. Cancer Res.* **14**, 6005–6013 (2008).
14. Sonoda, Y. *et al.* Formation of intracranial tumors by genetically modified human astrocytes defines four pathways critical in the development of human anaplastic astrocytoma. *Cancer Res.* **61**, 4956–4960 (2001).
15. Vivanco, I. *et al.* Identification of the JNK signaling pathway as a functional target of the tumor suppressor PTEN. *Cancer Cell* **11**, 555–569 (2007).
16. Sonoda, Y. *et al.* Akt pathway activation converts anaplastic astrocytoma to glioblastoma multiforme in a human astrocyte model of glioma. *Cancer Res.* **61**, 6674–6678 (2001).
17. Pieper, R. O. Defined human cellular systems in the study of glioma development. *Front. Biosci.* **8**, s19–s27 (2003).
18. Bibikova, M. *et al.* High density DNA methylation array with single CpG site resolution. *Genomics* **98**, 288–295 (2011).
19. Sandoval, J. *et al.* Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. *Epigenetics* **6**, 692–702 (2011).
20. Mutskov, V. & Felsenfeld, G. Silencing of transgene transcription precedes methylation of promoter DNA and histone H3 lysine 9. *EMBO J.* **23**, 138–149 (2004).
21. Bracken, A. P. *et al.* Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. *Genes Dev.* **20**, 1123–1136 (2006).
22. Furnari, F. B. *et al.* Malignant astrocytic glioma: genetics, biology, and paths to treatment. *Genes Dev.* **21**, 2683–2710 (2007).
23. Docherty, S. J. *et al.* Bisulfite-based epityping on pooled genomic DNA provides an accurate estimate of average group DNA methylation. *Epigenetics Chromatin* **2**, 3 (2009).
24. Hegi, M. E. *et al.* MGMT gene silencing and benefit from temozolomide in glioblastoma. *N. Engl. J. Med.* **352**, 997–1003 (2005).
25. Solomon, D. A. *et al.* Identification of p18^{INK4c} as a tumor suppressor gene in glioblastoma multiforme. *Cancer Res.* **68**, 2564–2569 (2008).
26. Etcheverry, A. *et al.* DNA methylation in glioblastoma: impact on gene expression and clinical outcome. *BMC Genomics* **11**, 701 (2010).
27. The Cancer Genome Atlas Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **455**, 1061–1068 (2008).
28. Lendahl, U., Zimmerman, L. B. & McKay, R. D. CNS stem cells express a new class of intermediate filament protein. *Cell* **60**, 585–595 (1990).
29. Xu, W. *et al.* Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of α -ketoglutarate-dependent dioxygenases. *Cancer Cell* **19**, 17–30 (2011).
30. He, Y. F. *et al.* Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **333**, 1303–1307 (2011).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Kayserian, K. Huberman, I. Dolgalev and S. Thomas for technical expertise. We thank C. Sawyers and E. Holland for helpful discussions. This grant was supported in part by the National Institutes of Health (R01CA154767-01) (T.A.C.), the National Cancer Institute (U54-CA143798) (I.K.M.), an Advanced Clinical Research Award in Glioma from the American Society of Clinical Oncology (I.K.M.), the Doris Duke Charitable Fund (I.K.M., T.A.C.), a National Brain Tumor Society Systems Biology Research Grant (I.K.M.), the MSKCC Brain Tumor Center (T.A.C.), the Louis Gerstner Foundation (T.A.C.), the STARR Cancer Consortium (T.A.C.) and the Sontag Foundation (T.A.C., I.K.M.).

Author Contributions T.A.C., S.T., A.G. and I.K.M. designed the experiments. S.T., A.G., F.F., D.R., A.H., L.A.W., C.C., E.Y., C.L., P.S.W., A.V., J.T.H., A.W.M.F. and L.G.T.M. performed the experiments. S.T., J.T.H., A.G., F.F., A.K., A.H., E.Y., A.V., P.S.W., C.B.T., T.A.C. and I.K.M. analysed the data. D.R., O.G., R.L. and I.K.M. contributed new reagents. T.A.C., S.T., I.K.M. and A.G. wrote the paper.

Author Information Data sets have been deposited in the Gene Expression Omnibus under accession number GSE30339. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to T.A.C. (chant@mskcc.org) or I.K.M. (mellingi@mskcc.org).

METHODS

Cell culture. Immortalized human astrocytes were a gift from R. O. Pieper (University of California, San Francisco) and were prepared as previously described¹⁴. Cells were cultured in Dulbecco's modified Eagle's medium (DMEM) plus 10% fetal bovine serum (FBS; Invitrogen). Expression of IDH was accomplished by cloning wild-type or mutant IDH1 (R132H) into the vector pLNCX2. These constructs were used to construct retroviruses used for infection of target cells. The retroviral packaging cell line GP-293 was seeded in 10-cm-diameter dishes and (at 30–50% confluency) was transfected using Lipofectamine (Invitrogen) with pVSV-G (Clontech) and pLNCX2-IDH1 wild type or IDH1 R132H. Retroviral particles were collected, filtered through a 0.45- μ m syringe filter and polybrene was added (8 μ g ml⁻¹ final concentration) to infect the human astrocytes for 12 h. Stable transfectants were selected with G418 and pooled populations of G418-resistant cells expressing either wild-type IDH1 or IDH1 R132H were confirmed by western blot analysis with anti-IDH1 antibody (rabbit anti-IDH1; Cell Signaling). All experiments were performed in duplicate.

Tumours. All tumours ($n = 81$) were obtained following surgical resection at the MSKCC as part of routine clinical care, and snap frozen. Tumours were obtained in accordance with Institutional Review Board policies at the MSKCC. Each sample was examined histologically by a neuropathologist. Before analysis, tumours were sectioned and microdissected. Genomic DNA or RNA was extracted using the DNeasy kit (Qiagen) or Triazol (Invitrogen) as per the manufacturer's instructions. Data from TCGA tumours ($n = 173$) are publicly available²⁷. For the LGG validation set, expression data sets of 115 patients with grade II and grade III gliomas were identified from the NCI Repository for Molecular Brain Neoplasia Data (Rembrandt; <http://rembrandt.nci.nih.gov>).

Sample preparation. DNA from wild-type IDH1, R132H IDH1 and parental astrocytes was extracted with the Puregene Cell and Tissue Kit (Qiagen) at various passages (passages 2, 5, 10, 15, 20, 25, 30, 40 and 50) and RNA was extracted with Trizol (Invitrogen) according to the manufacturer's directions. All experiments with the astrocytes were performed in duplicate, each with two corresponding technical (microarray) replicates. Genomic DNA and RNA from human tumours were extracted from frozen primary tumours for the methylation and expression studies. Frozen samples were snap frozen in liquid nitrogen and stored at -80°C . Each sample was examined histologically with haematoxylin-and-eosin-stained sections by a neuropathologist and representative sections were microdissected from the slides. Genomic DNA was extracted with the Qiagen DNeasy Blood and Tissue Kit using the manufacturer's instructions. RNA was extracted with Qiagen RNeasy Lipid Tissue Mini Kit using the manufacturer's instructions. Nucleic acid quality was determined with the Agilent 2100 Bioanalyzer.

Genomic analysis. Expression analysis of astrocytes and tumours was performed using the Affymetrix U133 2.0 microarray (Affymetrix). Genome-wide methylation analysis was performed using the Infinium HumanMethylation450 bead array (Illumina). Processing of the arrays was per the manufacturer's protocol. Methylation data were extracted using GenomeStudio software (Illumina). Methylation values for each site are expressed as a β value, representing a continuous measurement from 0 (completely unmethylated) to 1 (completely methylated). This value is based on following calculation: $\beta \text{ value} = (\text{signal intensity of methylation-detection probe}) / (\text{signal intensity of methylation-detection probe} + \text{signal intensity of non-methylation detection probe})$.

Data analysis. For methylation analysis, Illumina data were imported into Partek software. β Values were logit-transformed and adjusted for batch effects before analysis. Analysis of variance (ANOVA) with false discovery correction (FDR) was used to identify genes that were differentially methylated between the astrocytes expressing wild-type IDH1, mutant IDH1, and control astrocytes. Significant changes were defined as genes having an FDR-corrected P value < 0.05 . In human tumours, unsupervised consensus clustering of the β values was performed with K -means clustering ($K_{\text{max}} = 5$) with Euclidean distance and average linkage over 1,000 resampling iterations with random restart on the top 2% of the most variant probes (9,750 probes) using Gene Pattern v.2.0³¹. This identified an optimal number of $K = 2$ groups. This was repeated using unsupervised hierarchical clustering using Pearson dissimilarity. The cluster of samples that exhibited a large degree of hypermethylation was identified as CIMP+, and the remaining group CIMP-. ANOVA with FDR correction was used to identify genes that were differentially methylated between the CIMP groups. Significant changes were defined as genes having an FDR-corrected P value < 0.05 .

For gene expression analysis of astrocytes, Affymetrix CEL files were imported into the R statistical software (v.2.13.0; <http://www.R-project.org>). Normalization was performed with the AffyPLM package in BioConductor (v.2.4), using RMA background correction, quantile normalization, and the Tukey biweight summary method. Differential expression was detected using the limma package and P values were adjusted for multiple testing using the FDR approach. A probe set is considered differentially expressed if the FDR-adjusted P value < 0.05 . For gene

expression analysis of the human tumours, the Affymetrix data were imported into the Partek Genomics Suite (Partek) as Affymetrix CEL files. The data were RMA normalized and median-scaled for analysis. ANOVA followed by FDR was used to identify genes that were differentially expressed between the CIMP groups. To derive the 17-gene mutant IDH1 repression signature, we identified 605 unique genes that had either statistically significant hypermethylation at promoter-associated CpG islands and decreased gene expression, or had hypomethylation at promoter-associated CpG Islands and increased gene expression in CIMP+ versus CIMP- tumours. We identified common genes in a comparison of this gene set with that derived from mutant IDH1-expressing astrocytes versus wild type. Differential methylation in the cell lines was defined as an FDR-adjusted P value < 0.05 , and differential expression was defined as a P value < 0.05 with concordant fold change of at least 1.5 fold.

The 17-gene expression signature was used to predict CIMP in the Rembrandt data set. Unsupervised hierarchical clustering using Pearson dissimilarity identified two unique clusters that were categorized as 'predicted CIMP+' and 'predicted CIMP-'. The 17-gene expression signature was also used to identify subgroups from the TCGA GBM data set of 173 patients²⁷. Unsupervised consensus clustering using Pearson dissimilarity was performed on each of the subclasses identified by the TCGA to identify clusters³². Rembrandt data sets were obtained at <http://caintegrator-info.nci.nih.gov/rembrandt>.

Functional analysis of gene lists was performed using the PANTHER database and categories with adjusted P values (Benjamini-Hochberg) < 0.05 were considered as significantly over-represented in our gene lists³³. Concepts module mapping was performed as follows. The hypermethylation signature identified from our analysis of differentially methylated genes in IDH1 mutants compared to IDH1 wild-type was imported into OncoPrint (<http://www.oncoprint.org>) to identify associations with molecular concepts signatures derived from independent cancer profiling studies. Statistically significant concordances of our methylation gene signature with the pre-defined concepts were identified and Q value was calculated as previously described³⁴.

Methylation data of parental, wild-type IDH1-expressing astrocytes, and mutant IDH1-expressing astrocytes were clustered using self-organizing maps and visualized with the Gene Expression Dynamics Inspector (GEDI; v.2.1). For GEDI analysis, methylation data were normalized as a group across all passages and genotypes. Further hierarchical clustering (average-linkage) of GEDI map centroids was performed in R using the *hclust* library in the stats package. GSEA was performed using GSEA software v.2.0 and MSigDB database v.2.5. We assessed the significance of the curated gene sets (MSigDB collection c2) with the following parameters: number of permutations = 1,000 and permutation_type = phenotype, with an FDR Q -value cut-off of 5% (ref. 35).

G-CIMP was compared to MGMT methylation and MGMT expression in 52 LGG samples in the MSKCC cohort. We identified the Illumina 450K methylation probe ID (cg12981137) that corresponded to the MGMT MSP primer sequence as identified previously³⁶ and the Affymetrix probe ID (204880_at) that corresponds to MGMT expression.

Clinical and pathological characteristics between cohorts were compared using the χ^2 test. Overall survival was calculated from the date of surgery to death from any cause. Patients were censored at the time they were last known to be alive. Overall survival was assessed using the Kaplan-Meier method and the log-rank test was used for comparison between groups. Multivariate analysis was performed using a Cox proportional hazards model to assess the independent effect of prognostic variables on outcome, and using binary logistic regression to predict the probability of occurrence of CIMP+. An ROC curve was generated to graph the sensitivity and specificity of CIMP, MGMT methylation, and MGMT expression to predict survival ≥ 3 years. MGMT methylation and MGMT expression was considered continuous variables, and CIMP a categorical variable (defined by unsupervised hierarchical analysis as described above). Patients that were alive and had less than 3 years of follow-up were excluded from this analysis. Data was analysed using SPSS software (IBM SPSS statistics version 19.0).

Quantitative DNA methylation analysis using mass spectrometry. DNA methylation analysis was performed using the EpiTYPER system (Sequenom). The EpiTYPER assay is a tool for the detection and quantitative analysis of DNA methylation using base-specific cleavage of bisulphite-treated DNA and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS)³⁷. For primer sequences, target chromosomal sequence, and EpiTYPER-specific tags, see Supplementary Table 17. SpectroCHIPs were analysed using a Bruker Biflex III MALDI-TOF mass spectrometer (SpectroREADER, Sequenom). Results were analysed using the EpiTYPER Analyzer software, and manually inspected for spectra quality and peak quantification. CIMP positivity was defined as a mean methylated allelic frequency of $> 50\%$ or a twofold increase over normal breast tissue and the CIMP- state.

PCR amplification and sequencing. Exonic regions for the *IDH1* and *IDH2* genes (NCBI Human Genome Build 36.1) were broken into amplicons of 500 bp or less, and specific primers were designed using Primer3. Standard M13 tails were added to the primers to facilitate Sanger sequencing. PCR reactions were carried out in 384-well plates in a Duncan DT-24 water bath thermal cycler with 10 ng of whole-genome amplified DNA (REPLI-g Midi, Qiagen) as a template, using a touchdown PCR protocol with KAPA Fast HotStart (Kapa Biosystems). The touchdown PCR method consisted of: 1 cycle of 95 °C for 5 min; 3 cycles of 95 °C for 30 s, 64 °C for 15 s, 72 °C for 30 s; 3 cycles of 95 °C for 30 s, 62 °C for 15 s, 72 °C for 30 s; 3 cycles of 95 °C for 30 s, 60 °C for 15 s, 72 °C for 30 s; 37 cycles of 95 °C for 30 s, 58 °C for 15 s, 72 °C for 30 s; 1 cycle of 70 °C for 5 min. Templates were purified using AMPure (Agencourt Biosciences). The purified PCR reactions were split into two and sequenced bidirectionally with M13 forward and reverse primer and the Big Dye Terminator Kit v.3.1 (Applied Biosystems) at Agencourt Biosciences. Dye terminators were removed using the CleanSEQ kit (Agencourt Biosciences), and sequence reactions were run on ABI PRISM 3730xl sequencing apparatus (Applied Biosystems). Sanger sequencing of *IDH1* and *IDH2* produced an average coverage of 96.1% of coding sequence nucleotides across all samples.

Mutation detection. Passing reads were assembled against reference sequences, containing all coding exons including 5 kb upstream and downstream of the gene, using command line Consed 16.0³⁷. Assemblies were passed on to Polyphred 6.02b³⁸, which generated a list of putative candidate mutations, and to Polyscan 3.0³⁹, which generated a second list of putative mutations. The lists were merged together into a combined report, and the putative mutation calls were normalized to '+' genomic coordinates and annotated using the Genomic Mutation Consequence Calculator⁴⁰. The resulting list of annotated putative mutations was loaded into a Postgres database along with select assembly details for each mutation call (assembly position, coverage, and methods supporting mutation call). To reduce the number of false positives generated by the mutation detection software packages, only point mutations that were supported by at least one bi-directional read pair and at least one sample mutation called by Polyphred were considered, and only the putative mutations that were annotated as having non-synonymous coding effects, occurred within 1 bp of an exon boundary, or had a conservation score > 0.699 were included in the final candidate list. Indels were manually reviewed and included in the candidate list if found to hit an exon. All putative mutations were confirmed by a second PCR and sequencing reaction, in parallel with amplification and sequencing of matched normal tissue DNA.

ChIP. Cells were fixed with 1% formaldehyde for 10 min at room temperature (21 °C) and formaldehyde was inactivated by the addition of 125 mM glycine. ChIP assays were performed using a protocol recommended by the manufacturer of a commercially available ChIP assay kit (17-371, Millipore). Chromatin extracts were immunoprecipitated using anti-H3K9me3 (Ab8898, Abcam) or anti-H3K27me3 (07-449, Millipore) antibodies. After washing, ChIPed DNA was eluted from the beads and analysed on an Eppendorf Realplex using SYBR Green (Applied Biosystems). Relative occupancy values were calculated by determining ratios of the amount of immunoprecipitated DNA to that of the input sample (2% of total).

Flow cytometry and 5hmC assay. HEK 293T cells were transiently transfected with Flag-TET2 in pCMV6-ENTRY vector with Lipofectamine 2000 (GIBCO). For two-colour flow cytometry, 10⁶ cells were washed with ice-cold PBS, permeabilized and fixed using BD Cytoperm/Cytofix solution (BD, PharMingen), and incubated with anti-Flag (1:200, Sigma) and anti-5hmC (1:400, Active Motif #39770) antibodies for 30 min at room temperature. Cells were washed with PBS and incubated with secondary antibodies conjugated with Alexa Fluor 488 or Cy5 (Invitrogen) for 30 min in the dark. For single-colour flow cytometry, parental and *IDH1* mutant astrocytes were stained using anti-5hmC (1:400) followed by Alexa Fluor 488 secondary antibody. Cells were washed in PBS and analysed using the FACScan flow cytometer (Becton Dickinson). FACS data were analysed using FLOWJo Software (TreeStar).

Neurosphere assay. *IDH1*(R123H)-expressing astrocytes and parental controls were grown in media permissive of neural stem cell growth as previously described^{41,42}. Briefly, immortalized human astrocyte (IHA) cells stably expressing wild-type or R123H mutant *IDH1* at passage 15 were seeded in 6-well plates at 200,000 cells per well. The next day, proliferation medium (DMEM plus 10% FCS) was replaced with neural stem cell medium made from serum-free DMEM supplemented with B27 and N2 supplements (all from Invitrogen), bFGF, EGF and PDGFAA (all at 20 ng ml⁻¹, all from PeproTech). Medium was replaced every 2–3 days. Neurospheres were quantified using microscopy. Experiments were performed in triplicate.

31. Reich, M. *et al.* GenePattern 2.0. *Nature Genet.* **38**, 500–501 (2006).
32. Verhaak, R. G. *et al.* Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in *PDGFRA*, *IDH1*, *EGFR*, and *NF1*. *Cancer Cell* **17**, 98–110 (2010).
33. Thomas, P. D. *et al.* PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res.* **13**, 2129–2141 (2003).
34. Rhodes, D. R. *et al.* Molecular concepts analysis links tumors, pathways, mechanisms, and drugs. *Neoplasia* **9**, 443–454 (2007).
35. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
36. Vlassenbroeck, I. *et al.* Validation of real-time methylation-specific PCR to determine O⁶-methylguanine-DNA methyltransferase gene promoter methylation in glioma. *J. Mol. Diagn.* **10**, 332–337 (2008).
37. Gordon, D., Abajian, C. & Green, P. Consed: a graphical tool for sequence finishing. *Genome Res.* **8**, 195–202 (1998).
38. Nickerson, D. A., Tobe, V. O. & Taylor, S. L. PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. *Nucleic Acids Res.* **25**, 2745–2751 (1997).
39. Chen, K. *et al.* PolyScan: an automatic indel and SNP detection approach to the analysis of human resequencing data. *Genome Res.* **17**, 659–666 (2007).
40. Major, J. E. Genomic mutation consequence calculator. *Bioinformatics* **23**, 3091–3092 (2007).
41. Galli, R. *et al.* Isolation and characterization of tumorigenic, stem-like neural precursors from human glioblastoma. *Cancer Res.* **64**, 7011–7021 (2004).
42. De Filippis, L. *et al.* A novel, immortal, and multipotent human neural stem cell line generating functional neurons and oligodendrocytes. *Stem Cells* **25**, 2312–2321 (2007).

Transformation by the (*R*)-enantiomer of 2-hydroxyglutarate linked to EGLN activation

Peppi Koivunen^{1*}, Sungwoo Lee^{2*}, Christopher G. Duncan³, Giselle Lopez³, Gang Lu², Shakti Ramkissoon^{2,4,5}, Julie A. Losman², Päivi Joensuu⁶, Ulrich Bergmann⁷, Stefan Gross⁸, Jeremy Travins⁸, Samuel Weiss⁹, Ryan Looper¹⁰, Keith L. Ligon^{2,4,5,11}, Roel G. W. Verhaak¹², Hai Yan³ & William G. Kaelin Jr^{2,13}

The identification of succinate dehydrogenase (SDH), fumarate hydratase (FH) and isocitrate dehydrogenase (IDH) mutations in human cancers has rekindled the idea that altered cellular metabolism can transform cells. Inactivating SDH and FH mutations cause the accumulation of succinate and fumarate, respectively, which can inhibit 2-oxoglutarate (2-OG)-dependent enzymes, including the EGLN prolyl 4-hydroxylases that mark the hypoxia inducible factor (HIF) transcription factor for polyubiquitylation and proteasomal degradation¹. Inappropriate HIF activation is suspected of contributing to the pathogenesis of SDH-defective and FH-defective tumours but can suppress tumour growth in some other contexts. IDH1 and IDH2, which catalyse the interconversion of isocitrate and 2-OG, are frequently mutated in human brain tumours and leukaemias. The resulting mutants have the neomorphic ability to convert 2-OG to the (*R*)-enantiomer of 2-hydroxyglutarate ((*R*)-2HG)^{2,3}. Here we show that (*R*)-2HG, but not (*S*)-2HG, stimulates EGLN activity, leading to diminished HIF levels, which enhances the proliferation and soft agar growth of human astrocytes. These findings define an enantiomer-specific mechanism by which the (*R*)-2HG that accumulates in IDH mutant brain tumours promotes transformation and provide a justification for exploring EGLN inhibition as a potential treatment strategy.

To study the role of IDH mutations in brain tumours, we stably infected immortalized human astrocytes with retroviral vectors encoding haemagglutinin (HA)-tagged versions of wild-type IDH1, a tumour-derived mutant (IDH1 R132H)^{2,3}, or an IDH1 R132H variant in which three conserved aspartic acid residues within the IDH1 catalytic domain were replaced with asparagines (R132H/3DN) (Fig. 1a and Supplementary Fig. 1). As expected, (*R*)-2HG levels, but not (*S*)-2HG levels, were markedly increased in the cells producing IDH1 R132H but not in cells producing the R132H/3DN variant (Fig. 1b and Supplementary Fig. 2). In multiple independent experiments the IDH1 R132H cells acquired a proliferative advantage relative to cells producing the other versions of IDH1 beginning around passage 14, manifested as increased proliferation at confluence (Fig. 1c) and the ability to form macroscopic colonies in soft agar (Fig. 1d, e).

Consistent with recent reports, we found that both (*R*)-2HG and (*S*)-2HG inhibit a number of 2-OG-dependent enzymes *in vitro*^{4–6}, including the collagen prolyl 4-hydroxylases, the TET1 and TET2 methyl cytosine hydroxylases, the HIF asparaginyl hydroxylase FIH1 (also known as HIF1AN) and the JMJD2D (also known as KDM4D) histone demethylase, with (*S*)-2HG being a more potent inhibitor than (*R*)-2HG (Supplementary Fig. 3 and data not shown). (*S*)-2HG was

also a micromolar to low millimolar inhibitor of the three mammalian HIF prolyl 4-hydroxylases (EGLN1, EGLN2 and EGLN3) under standard assay conditions, which included 10 μ M 2-OG (Supplementary Fig. 4). In contrast, (*R*)-2HG was not an effective EGLN inhibitor (half-maximum inhibitory concentration (IC₅₀) values > 5 mM; Supplementary Fig. 4). Moreover, we discovered unexpectedly that (*R*)-2HG, but not (*S*)-2HG, promoted EGLN1 and EGLN2 activity, and to a lesser extent EGLN3 activity, at tumour-relevant concentrations (low mM) in reactions that lacked exogenous 2-OG (Fig. 2a, c, d and Supplementary Fig. 5a, b). This was specific because (*R*)-2HG did not promote collagen prolyl 4-hydroxylase activity (Fig. 2b) or JMJD2D histone demethylase activity (data not shown). Similar results were obtained with EGLN1 purified from either insect cells or *Escherichia coli* and with a heat-inactivated HIF polypeptide substrate (Supplementary Fig. 6), making it unlikely that the ability of (*R*)-2HG to promote EGLN activity required a contaminating enzyme.

To determine how (*R*)-2HG might promote EGLN activity, we monitored the EGLN1 prolyl 4-hydroxylase reaction using liquid

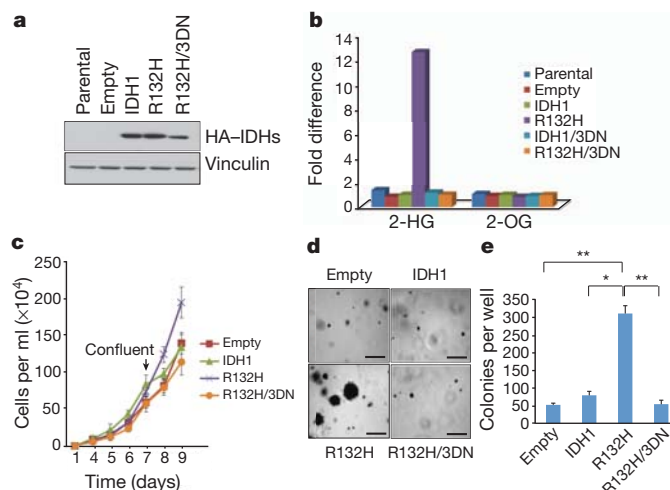


Figure 1 | Oncogenic properties of IDH1 R132H. **a**, **b**, Anti-HA immunoblot (**a**) and LC-MS analysis (**b**) of immortalized human astrocytes (passage 4) infected with retroviruses encoding HA-tagged versions of the indicated IDH1 variants. **c**, **d**, *In vitro* proliferation under standard culture conditions (**c**) or in soft agar (**d**) (passage 23). Scale bars, 0.5 mm. **e**, Number of macroscopic soft agar colonies in **d** (**P* < 0.01, ***P* < 0.005). Error bars show standard deviation (s.d.); *n* = 3.

¹Biocenter Oulu, Department of Medical Biochemistry and Molecular Biology, Oulu Center for Cell-Matrix Research, University of Oulu, FIN-90014 Oulu, Finland. ²Department of Medical Oncology, Dana-Farber Cancer Institute and Brigham and Women's Hospital, Boston, Massachusetts 02215, USA. ³The Preston Robert Tisch Brain Tumor Center at Duke, The Pediatric Brain Tumor Foundation Institute & The Department of Pathology, Duke University Medical Center, Durham, North Carolina 27710, USA. ⁴Department of Pathology, Brigham and Women's Hospital, Boston, Massachusetts 02215, USA. ⁵Department of Pathology, Harvard Medical School, Boston, Massachusetts 02215, USA. ⁶Department of Chemistry, University of Oulu, FIN-90014 Oulu, Finland. ⁷Biocenter Oulu, Mass Spectrometry Core Facility, Department of Biochemistry, University of Oulu, FIN-90014 Oulu, Finland. ⁸Agiros Pharmaceuticals, Cambridge, Massachusetts 02139, USA. ⁹Hotchkiss Brain Institute, Departments of Cell Biology and Anatomy, University of Calgary Faculty of Medicine, Calgary, Alberta T2N 4N1, Canada. ¹⁰Department of Chemistry, University of Utah, Salt Lake City, Utah 84112, USA. ¹¹Department of Pathology, Children's Hospital Boston, Boston, Massachusetts 02215, USA. ¹²Department of Bioinformatics and Computational Biology, University of Texas MD Anderson Cancer Center, Houston, Texas 77030, USA. ¹³Howard Hughes Medical Institute, Chevy Chase, Maryland 20815, USA.

*These authors contributed equally to this work.

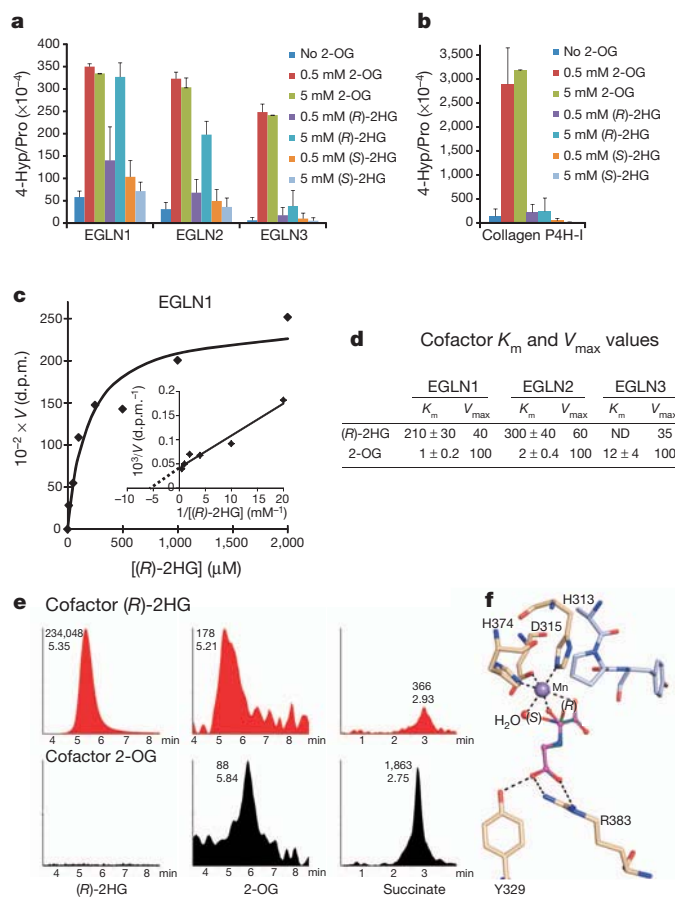


Figure 2 | (R)-2HG can serve as an EGLN cosubstrate. **a, b**, *In vitro* prolyl 4-hydroxylation assays conducted with recombinant EGLN proteins (**a**) and collagen P4H-I (**b**) in the presence of the indicated amounts of 2-OG or 2HG. 4-Hyp/Pro, ratio of 4-hydroxyproline to proline. L-[2,3,4,5- 3 H]proline-labelled HIF-1 α oxygen-dependent degradation domain (ODDD) (**a**) and [14 C]proline-labelled procollagen (**b**) were used as substrates. Enzymes were produced in insect cells using baculoviruses and affinity purified. Error bars, s.d.; $n = 3$ –4. **c, d**, K_m (in μ M) and maximum enzyme velocity (V_{max} ; percentage of that obtained with 2-OG) values for (R)-2HG for EGLN family members. K_m and V_{max} values for 2-OG²¹ are included for comparison. d.p.m., disintegrations per minute. N.D., not defined. **e**, LC-MS analysis of succinate, 2-OG and (R)-2HG from enzymatic reactions with EGLN1, HIF-1 α ODDD polypeptide and either 5 mM (R)-2HG (red) or 80 μ M 2-OG (black) as cofactors. Numbers next to each peak indicate elution times (bottom) and peak areas (top). No peaks above background were detected in samples in which 2-OG and (R)-2HG were both omitted (data not shown). **f**, Model of (R)-2HG (green) and (S)-2HG (cyan) bound to the active site of EGLN1. N-oxalylglycine (magenta) bound in the original structure²⁹ is shown for comparison. The active-site water molecule, which has been shown to be the O₂-binding site³⁰, is shown in red and the peptide substrate in light blue. Hydrogen bonds are indicated by dash lines.

chromatography-mass spectrometry (LC-MS). 2-OG and succinate were detected when catalytically active EGLN1 was incubated with 5 mM (R)-2HG and a recombinant HIF-1 α polypeptide, suggesting that EGLN1 can oxidize (R)-2HG to 2-OG, which is then decarboxylated to succinate during the hydroxylation reaction (Fig. 2e and Supplementary Fig. 5c). In support of this model, 13 C-labelled succinate was generated in EGLN hydroxylation assays that contained uniformly labelled 13 C-(R)-2HG (Supplementary Fig. 5d).

Consistent with the idea that (R)-2HG can substitute for 2-OG as a cosubstrate, the addition of increasing amounts of (R)-2HG to EGLN1 assays containing 10 μ M 2-oxo-[1- 14 C]glutamate progressively decreased the release of 14 CO₂ without decreasing the prolyl hydroxylation of HIF-1 α , even at concentrations as high as 100 mM (Supplementary Fig. 7). Modelling of 2HG bound to the active site of EGLN1 predicts that binding of the (S)-enantiomer, but not the (R)-enantiomer, would

prevent the subsequent recruitment of oxygen to the active site, perhaps accounting for the qualitatively different effects of the two enantiomers on EGLN activity (Fig. 2f and Supplementary Fig. 5e).

To investigate whether these findings were relevant *in vivo* we examined HIF levels in IDH1 mutant cells. In keeping with our biochemical results, HIF-1 α and HIF-2 α protein levels were reproducibly lower in mid-passage (passage 9–15) human astrocytes producing IDH1 R132H relative to control astrocytes (Fig. 3a), due at least partly to increased HIF-1 α and HIF-2 α hydroxylation and diminished protein stability (Supplementary Figs 8 and 9), and were associated with lower levels of the HIF-responsive messenger RNAs encoding *VEGF*, *GLUT1* and *PDK1* (Supplementary Fig. 10a). Oxygen consumption and reactive oxygen species production, which can also affect HIF levels and the HIF response, were not measurably altered in the IDH1 mutant cells (Supplementary Fig. 11). IDH1(R132H)-expressing cells were relatively resistant to the 2-OG competitive antagonist dimethylxalylglycine (DMOG), but not to the iron chelator deferoxamine (DFO) (Fig. 3b), consistent with (R)-2HG acting as a 2-OG agonist in intact cells. In later passages (more than passage 20), HIF levels in IDH1(R132H)-expressing immortalized human astrocytes began to normalize despite persistent production of the exogenous IDH1 protein and (R)-2HG (Supplementary Fig. 12), possibly due to adaptive HIF-responsive feedback loops such as those involving EGLN3 and microRNA miR-155 (refs 1, 7). These cells, however, retained the ability to form colonies in soft agar (data not shown) and remained addicted to EGLN activity (see later).

Similarly, the induction of HIF-1 α and HIF-responsive mRNAs by hypoxia was diminished in two independent cell lines derived from two IDH1 R132H, 1p/19q-codeleted oligodendrogliomas compared to a control IDH1 wild-type, 1p/19q-codeleted oligodendroglioma line that had been generated in a similar fashion (Fig. 3c and Supplementary Fig. 10b). Although these three cell lines have similar growth kinetics *in vitro* (data not shown) they are not isogenic. We therefore also tested two HCT116 colorectal cancer cell sublines wherein the R132H mutation was introduced into the endogenous *IDH1* locus by homologous recombination. These sublines also showed a diminished HIF response compared to wild-type cells unless EGLN was pharmacologically (DFO) or genetically (short hairpin (sh)RNA) inactivated (Fig. 3d, e and Supplementary Fig. 13).

Finally, we asked whether IDH mutational status influenced HIF activity in primary patient astrocytoma samples using the TCGA expression data set⁸ and a previously defined HIF-responsive gene expression signature⁹. The HIF signature was diminished in proneural tumours—which is the subtype most often associated with IDH mutations⁸—relative to other gene-expression-defined subtypes of brain cancer (Supplementary Fig. 14a, b) and, notably, was diminished in IDH mutant proneural tumours relative to wild-type proneural tumours (Fig. 3f and Supplementary Fig. 14c). Similar results were obtained with other previously published HIF gene sets and with a manually curated data set (Supplementary Fig. 14a, d).

Downregulating HIF-1 α with three independent shRNAs promoted soft agar growth by immortalized human astrocytes after approximately 15 passages (Fig. 4a, b and Supplementary Fig. 15), as did overproduction of wild-type, but not catalytic-defective, EGLN1 (Fig. 4c, d and Supplementary Fig. 16). Conversely, downregulation of EGLN1 with multiple shRNAs inhibited the proliferation of late passage IDH1 R132H cells (Fig. 4e, f and Supplementary Fig. 17) unless HIF-1 α was concurrently ablated (Fig. 4g, h and Supplementary Fig. 18).

Collectively, these data suggest that EGLN activation by (R)-2HG, and subsequent downregulation of HIF-1 α , contributes to the pathogenesis of IDH mutant gliomas. Our data do not, however, exclude the possibility that (R)-2HG has additional targets, including TET2 and JmJC-containing histone demethylases, which contribute to its ability to transform cells^{4–6}. Indeed, we found that downregulation of TET2 in human astrocytes also promotes soft agar growth (Supplementary

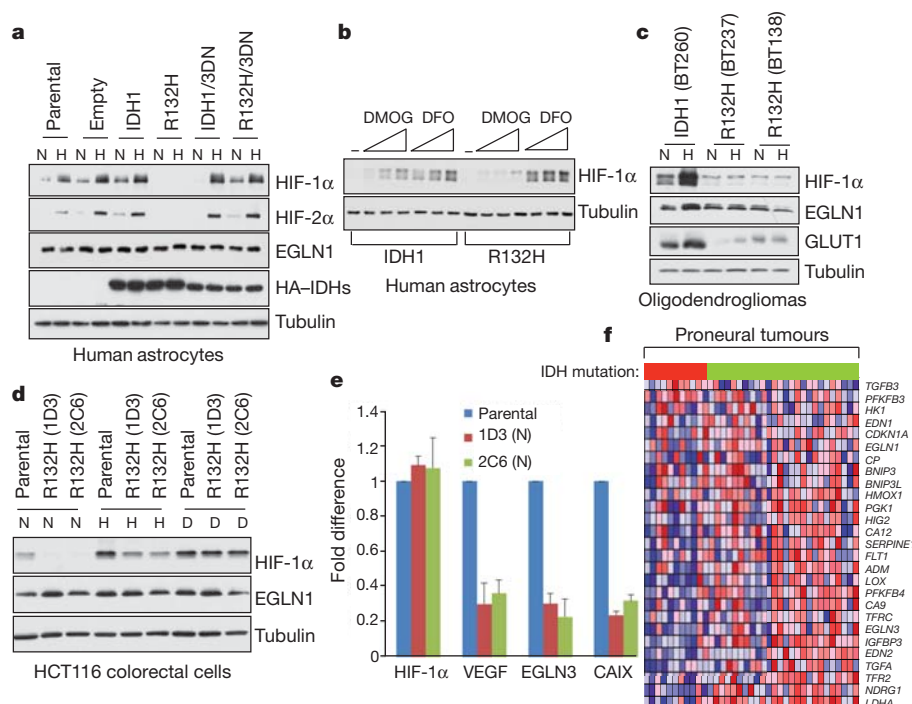


Figure 3 | HIF activity is diminished in IDH mutant cells. **a–d**, Immunoblot analysis of immortalized human astrocytes (passage 10) (**a**, **b**), oligodendroglioma cells (**c**) and HCT116 colorectal cells (**d**) expressing the indicated IDH1 variants grown under 21% (N) or 7.5% (H) oxygen for 24 h before lysis or under 21% oxygen in the presence of DFO (D). Names of cell lines are indicated in brackets. In **b**, cells grown under 21% oxygen were treated

with increasing amounts of DMOG or DFO for 16 h before lysis. **e**, Quantitative real-time PCR analysis of cells in **d** under normoxic conditions. Error bars, s.d. $n = 3$. **f**, Heat map depicting expression of HIF target genes (blue, lower expression; red, higher expression) in proneural tumours clustered based on IDH status (in the horizontal bar above the matrix, red indicates mutant; green, wild type).

Fig. 19). Nonetheless, our finding that (R)-2HG and (S)-2HG are qualitatively different with respect to EGLN might explain the apparent selection for (R)-2HG in adult tumours, despite the fact that (S)-2HG is a more potent inhibitor of most of the 2-OG-dependent enzymes tested so far. It should be noted, however, that (S)-2HG has been linked to neurological abnormalities and brain tumours in children and young adults with germline (S)-2-HG dehydrogenase mutations^{10–12}. The pathogenesis of (R)-2HG-driven tumours linked to somatic IDH mutations in adults and (S)-2HG-driven tumours linked to inborn errors of metabolism conceivably differ, with the latter possibly reflecting the perturbation of one or more neurodevelopmental programs during embryogenesis.

Nor are our data incompatible with the finding that HIF-1α protein levels are increased in IDH mutant tumours relative to normal brain¹³. Our data simply suggest that (R)-2HG quantitatively shifts the dose-response linking HIF activation to hypoxia, leading to a blunted HIF response for a given level of hypoxia. In support of this idea, HIF elevation in IDH mutant tumours is usually confined to areas of necrosis and presumed severe hypoxia¹⁴.

Although HIF is typically viewed as an oncoprotein it can behave as a tumour suppressor in embryonic stem cells^{15,16}, leukaemic cells¹⁷ and brain tumour cells^{18,19}. For example, HIF-1α has been shown to score as an oncoprotein when transformed murine astrocytes are grown subcutaneously but as a tumour suppressor when such cells are grown orthotopically¹⁹. This finding, together with our results, raises the possibility that pharmacological inhibition of EGLN activity (and the resulting increase in HIF activity) would impair the growth of IDH1 mutant tumours. A caveat, however, is that IDH mutant tumours tend to be relatively indolent²⁰. It is possible that low levels of HIF-1α, although promoting some aspects of transformation, simultaneously suppress other hallmarks of cancer required for aggressive behaviour (such as angiogenesis).

The published 2-OG Michaelis constant (K_m) values for the EGLN family members are below the estimated intracellular 2-OG concentration^{21,22}, suggesting that 2-OG should not be limiting for EGLN activity and that EGLN activity would not be enhanced further by (R)-2HG. A caveat is that a considerable amount of intracellular 2-OG appears to be sequestered in mitochondria and might also be bound by other 2-OG-dependent enzymes²². Moreover, these 2-OG K_m values were determined under idealized conditions with purified enzymes and substrates in the absence of endogenous inhibitors such as reactive oxygen species, nitric oxide and 2-OG competitive molecules such as succinate and fumarate¹. Studies in model organisms suggest that many metabolic enzymes are saturated *in vivo* with a mixture of substrate and competitive inhibitors and thus are sensitive to changes in substrate concentrations that are far above their nominal K_m values^{23,24}. We confirmed that succinate and fumarate increase the 2-OG requirement for the EGLN reaction and that their inhibitory activity, like that of DMOG (Fig. 3b) and its active derivative 2-oxalyl glycine, is blunted in the presence of (R)-2HG (Supplementary Fig. 20).

We have not yet formally proven that (R)-2HG is sufficient to downregulate HIF in intact cells. It is possible, for example, that additional metabolic changes in IDH mutant cells sensitize them to the HIF modulatory effects of (R)-2HG²⁵. Another, potentially related, observation is that both downregulation of HIF and transformation by mutant IDH in immortalized human astrocytes, although highly reproducible, was only noted after multiple passages. HIF activates a number of genes, including genes that participate in feedback regulation of the HIF response and genes that modify chromatin structure¹. It is possible that modulation of the HIF response over time, perhaps in conjunction with alterations in other enzymes affected by 2-HG, leads to epigenetic changes that ultimately are responsible for transformation. If so, it will be important to determine the degree to which such changes are reversible.

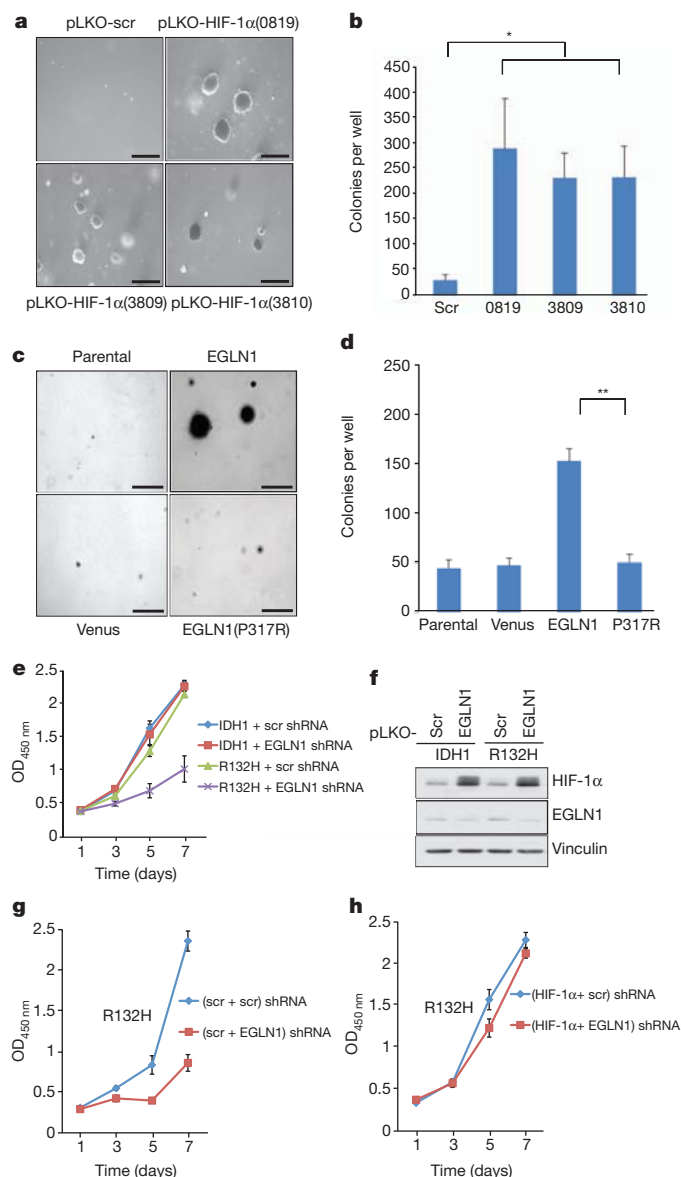


Figure 4 | Decreased HIF activity contributes to transformation by mutant IDH. **a, c**, Soft agar colony formation by human astrocytes after stable infection with lentiviruses encoding the indicated HIF-1 α shRNAs or scrambled (scr) shRNA (**a**) or lentiviruses encoding EGLN1, EGLN1 (P317R), or venus fluorescent protein (**c**; RNAi Consortium shRNA clone identifiers are indicated in brackets). Scale bars, 0.5 mm. **b, d**, Number of macroscopic soft agar colonies in **a** and **c**, respectively. * $P < 0.05$, ** $P < 0.005$. Error bars, s.d. $n = 3$. **e, f**, Proliferation (**e**) and immunoblot analysis (**f**) of human astrocytes expressing wild-type or R132H IDH1 and an shRNA against EGLN1 (10578) (or scrambled control). **g, h**, Proliferation of human astrocytes expressing IDH1 R132H and shRNAs against HIF-1 α (0819), EGLN1 (10578), or both.

METHODS SUMMARY

NHAs immortalized with E6/E7/hTERT have been described elsewhere²⁶ and were subsequently infected with IDH retroviruses. IDH1 R132H HCT116 colorectal cancer cells were made by homologous recombination as previously described²⁷. Anaplastic oligodendroglioma cell lines were derived from resection material obtained at surgeries performed at the Brigham and Women's Hospital. Lentiviral shRNA vectors were from the Broad Institute. Immunoblots were performed with anti-HIF-1 α monoclonal antibody (BD Transduction Laboratories), anti-HIF-2 α polyclonal antibody (NB 100-122, Novus), anti-Flag monoclonal antibody (M2, Sigma-Aldrich), monoclonal anti-HA (HA-11, Covance Research Product), anti-EGLN1 monoclonal antibody (D31E11, Cell Signaling), anti-GLUT1 polyclonal antibody (NB300-666, Novus Biologicals), monoclonal anti-tubulin (B-512, Sigma-Aldrich) or anti-vinculin monoclonal antibody (V9131, Sigma-Aldrich). Metabolite levels were determined by LC-MS as previously

described². Enzyme activity assays were performed with affinity-purified human EGLN1, 2 and 3, collagen P4H-I and FIH, and murine Tet1 and Tet2 produced in insect cells using baculoviral vectors. (R)-2HG and (S)-2HG were from Sigma. (R)-2HG was free of contaminating 2-OG as determined by LC-MS under conditions that could detect 0.37 μ M exogenous 2-OG in 5 mM (R)-2HG (data not shown). Human brain tumour mRNA expression data were obtained from The Cancer Genome Atlas and processed as described^{8,28}. Cell proliferation assays were performed using XTT assays (Cell Proliferation Kit II, Roche) according to the manufacturer's instructions or by direct counting of viable cells using an automated cell counter (Invitrogen). For soft agar assays ~8,000 cells were suspended in a top layer of 0.4% soft agar (SeaPlaque Agarose, BMA products) and plated on a bottom layer of 1% soft agar containing DMEM supplemented with 10% fetal bovine serum. For real-time qPCR analysis total RNAs were first extracted with Trizol reagent (Invitrogen). cDNA synthesis and PCR amplification were performed with Superscript One-Step RT-PCR (Invitrogen).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 26 May 2011; accepted 26 January 2012.

Published online 15 February 2012.

- Majmudar, A. J., Wong, W. J. & Simon, M. C. Hypoxia-inducible factors and the response to hypoxic stress. *Mol. Cell* **40**, 294–309 (2010).
- Dang, L. *et al.* Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature* **462**, 739–744 (2009).
- Jin, G. *et al.* 2-Hydroxyglutarate production, but not dominant negative function, is conferred by glioma-derived NADP⁺-dependent isocitrate dehydrogenase mutations. *PLoS ONE* **6**, e16812 (2011).
- Figuroa, M. E. *et al.* Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. *Cancer Cell* **18**, 553–567 (2010).
- Xu, W. *et al.* Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of α -ketoglutarate-dependent dioxygenases. *Cancer Cell* **19**, 17–30 (2011).
- Chowdhury, R. *et al.* The oncometabolite 2-hydroxyglutarate inhibits histone lysine demethylases. *EMBO Rep.* **12**, 463–469 (2011).
- Bruning, U. *et al.* MicroRNA-155 promotes resolution of hypoxia-inducible factor-1 α activity during prolonged hypoxia. *Mol. Cell. Biol.* **31**, 4087–4096 (2011).
- Verhaak, R. G. *et al.* Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in *PDGFRA*, *IDH1*, *EGFR*, and *NF1*. *Cancer Cell* **17**, 98–110 (2010).
- Nickols, N. G., Jacobs, C. S., Farkas, M. E. & Dervan, P. B. Modulating hypoxia-inducible transcription by disrupting the HIF-1-DNA interface. *ACS Chem. Biol.* **2**, 561–571 (2007).
- Aghili, M., Zahedi, F. & Rafiee, E. Hydroxyglutaric aciduria and malignant brain tumor: a case report and literature review. *J. Neurooncol.* **91**, 233–236 (2009).
- Moroni, I. *et al.* L-2-hydroxyglutaric aciduria and brain malignant tumors: a predisposing condition? *Neurology* **62**, 1882–1884 (2004).
- Özışık, P. A., Akalan, N., Palaoglu, S. & Topcu, M. Medulloblastoma in a child with the metabolic disease L-2-hydroxyglutaric aciduria. *Pediatr. Neurosurg.* **37**, 22–26 (2002).
- Zhao, S. *et al.* Glioma-derived mutations in IDH1 dominantly inhibit IDH1 catalytic activity and induce HIF-1 α . *Science* **324**, 261–265 (2009).
- Williams, S. C. *et al.* R132H-mutation of isocitrate dehydrogenase-1 is not sufficient for HIF-1 α upregulation in adult glioma. *Acta Neuropathol.* **121**, 279–281 (2011).
- Carmeliet, P. *et al.* Role of HIF-1 α in hypoxia-mediated apoptosis, cell proliferation and tumour angiogenesis. *Nature* **394**, 485–490 (1998).
- Mack, F. A. *et al.* Loss of pVHL is sufficient to cause HIF dysregulation in primary cells but does not promote tumor growth. *Cancer Cell* **3**, 75–88 (2003).
- Song, L. P. *et al.* Hypoxia-inducible factor-1 α -induced differentiation of myeloid leukemic cells is its transcriptional activity independent. *Oncogene* **27**, 519–527 (2008).
- Acker, T. *et al.* Genetic evidence for a tumor suppressor role of HIF-2 α . *Cancer Cell* **8**, 131–141 (2005).
- Blouw, B. *et al.* The hypoxic response of tumors is dependent on their microenvironment. *Cancer Cell* **4**, 133–146 (2003).
- Christensen, B. C. *et al.* DNA methylation, isocitrate dehydrogenase mutation, and survival in glioma. *J. Natl. Cancer Inst.* **103**, 143–153 (2011).
- Koivunen, P. *et al.* Inhibition of hypoxia-inducible factor (HIF) hydroxylases by citric acid cycle intermediates: possible links between cell metabolism and stabilization of HIF. *J. Biol. Chem.* **282**, 4524–4532 (2007).
- Pritchard, J. B. Intracellular α -ketoglutarate controls the efficacy of renal organic anion transport. *J. Pharmacol. Exp. Ther.* **274**, 1278–1284 (1995).
- Bennett, B. D. *et al.* Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. *Nature Chem. Biol.* **5**, 593–599 (2009).
- Yuan, J. *et al.* Metabolomics-driven quantitative analysis of ammonia assimilation in *E. coli*. *Mol. Syst. Biol.* **5**, 302 (2009).
- Reitman, Z. J. *et al.* Profiling the effects of isocitrate dehydrogenase 1 and 2 mutations on the cellular metabolome. *Proc. Natl Acad. Sci. USA* **108**, 3270–3275 (2011).

26. Sonoda, Y. *et al.* Formation of intracranial tumors by genetically modified human astrocytes defines four pathways critical in the development of human anaplastic astrocytoma. *Cancer Res.* **61**, 4956–4960 (2001).
27. Rago, C., Vogelstein, B. & Bunz, F. Genetic knockouts and knockins in human somatic cells. *Nature Protocols* **2**, 2734–2746 (2007).
28. Wang, X. V., Verhaak, R. G., Purdom, E., Spellman, P. T. & Speed, T. P. Unifying gene expression measures from multiple platforms using factor analysis. *PLoS ONE* **6**, e17691 (2011).
29. Chowdhury, R. *et al.* Structural basis for binding of hypoxia-inducible factor to the oxygen-sensing prolyl hydroxylases. *Structure* **17**, 981–989 (2009).
30. Koski, M. K. *et al.* The active site of an algal prolyl 4-hydroxylase has a large structural plasticity. *J. Biol. Chem.* **282**, 37112–37123 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. P. Hausinger and J. D. Rabinowitz for helpful suggestions and critical reading of the manuscript, C. Schofield and Y. Zhang for reagents, S. Chen and Y. Shi for JMJD2D assays, K. Koski for modelling and T. Aatsinki and E. Lehtimäki for technical assistance. W.G.K. is a Doris Duke Distinguished Clinical Scholar and a Howard Hughes Medical Institute (HHMI) Investigator. Supported by the

National Institutes of Health (W.G.K.), HHMI (W.G.K.), Doris Duke Foundation (W.G.K.), Academy of Finland Grants 120156, 140765 and 218129 (P.K.) and S. Juselius Foundation (P.K.).

Author Contributions P.K., S.L. and W.G.K. initiated the project, analysed the data and wrote the manuscript. P.K. performed the *in vitro* enzyme assays. S.L. generated astrocyte cell lines stably expressing various IDH1 proteins. C.G.D., G. Lo. and H.Y. generated the HCT116 subclones. S.R., K.L.L. and S.W. provided oligodendroglioma cell lines. G. Lu generated and validated the reporter plasmids encoding HIF-1 α -luciferase fusion proteins. P.J., U.B. and S.G. performed the LC-MS analysis. J.T. synthesized ¹³C-(R)-2HG and R.L. synthesized and purified different 2-OG and 2-HG derivatives. R.G.W.V. performed the bioinformatics. P.K. and S.L. performed all other experiments with the help of G. Lu, J. A. L. and P.J. All the authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to W.G.K. (william_kaelin@dfci.harvard.edu).

METHODS

Cell lines. The NHA cell line immortalized with E6/E7/hTERT has been described elsewhere²⁶ and was maintained in DMEM containing 10% fetal bovine serum (FBS) and 1% penicillin/streptomycin in the presence of 10% CO₂ at 37 °C. Following retroviral infection, cells were maintained in the presence of hygromycin (100 µg ml⁻¹).

Introduction of the IDH1 R132H mutation into HCT116 colon cancer cell lines by homologous recombination was as previously described²⁷. Briefly, targeting constructs were designed using the pSEPT rAAV shuttle vector³¹. Homology arms for the targeting vector were PCR amplified from HCT116 genomic DNA using Platinum Taq HiFi polymerase (Invitrogen). The R132H hotspot mutation was introduced in the targeting construct using the Quickchange II site-directed mutagenesis kit (Stratagene). An infectious rAAV stock harbouring the targeting sequence was generated and applied to parental HCT116 cells as previously described³², and clones were selected in 0.5 mg ml⁻¹ geneticin (Invitrogen). Excision of the selectable element was achieved with an adenovirus encoding Cre recombinase (Vector Biolabs). Genomic DNA and total RNA were isolated from cells using QIAmp DNA Blood Kit and RNeasy kit (Qiagen). First strand cDNA was synthesized using iScript cDNA Synthesis Kit (BioRad). Successful homologous recombination and Cre-mediated excision were verified using PCR-based assays and by direct sequencing of genomic DNA and cDNA.

Anaplastic oligodendroglioma cell lines expressing wild-type IDH1 (BT260) or IDH1 R132H (BT237 and BT138) were derived from surgical resection material acquired from patients undergoing surgery at the Brigham and Women's Hospital on an Institutional Review Board approved protocol. Briefly, tumour resection samples were mechanically dissociated and tumospheres were established and propagated in Human NeuroCult NS-A Basal media (StemCell Technologies) supplemented with EGF, FGFb and heparin sulphate. All lines were obtained from the DF/BWCC Living Tissue Bank and confirmed to be derived from recurrent and progressive anaplastic oligodendrogliomas, World Health Organization grade III, at the time of cell-line isolation, and to have chromosome 1p/19q codeletion. The presence of IDH R132H mutation was confirmed by mutant specific antibody staining via immunohistochemistry and direct DNA sequencing.

Vectors. Human IDH1 cDNAs (wild type and R132H) were subcloned as BamHI-EcoRI fragments into pBabe-HA-hygro. Aspartic acid residues 273, 275 and 279 were mutated to asparagines (N) by site-directed mutagenesis and confirmed by DNA sequencing. *EGLN1*, *EGLN2* and *EGLN3* cDNAs were subcloned into pLenti6-Flag expression vector after restriction with XbaI and EcoRI. pLenti6-Flag-EGLN1 P317R³³ was prepared by QuikChange Mutagenesis (Stratagene) using pLenti6-Flag EGLN1 as a template and the following oligonucleotides: 5'-GTACGTCATGTTGATAATCGAAATGGAGATGGAAGATCTG-3' and 5'-CACATGTTCCATCTCCATTTCGATTATCAACATGACGTAC-3'. The entire EGLN1 coding region was sequenced to verify its authenticity.

Lentiviral (pLKO.1) HIF-1α shRNA vectors (TRCN0000003810, target sequence: 5'-GTGATGAAAGAATTACCGAAT-3'; TRCN0000010819, target sequence: 5'-TGCTCTTTGTGGTTGGATCTA-3'; TRCN0000003809, target sequence: 5'-CCAGTTATGATTGTGAAGTTA-3'), EGLN1 (PHD2) shRNA vectors (TRCN0000001042, target sequence: 5'-CTGTTATCTAGCTGAGTT CAT-3'; TRCN0000001043, target sequence: 5'-GACGACCTGATACGCCACT GT-3'; TRCN00000010578, target sequence: 5'-TGCACGACACCGGGAAGT TCA-3') and TET2 shRNA vectors (TRCN0000122172 (122), target sequence: 5'-GCGTTTATCCAGAAATTAGCAA-3'; TRCN0000144344 (145), target sequence: 5'-CCTTATAGTCAGACCATGAAA-3') were obtained from the Broad Institute TRC shRNA library. pLKO.1 shRNA with target sequence 5'-GCAAG CTGACCTGAAGTTCAT-3' was used as negative control shRNA.

Immunoblot analysis. Cells extracts were prepared with 1× lysis buffer (50 mM Tris (pH 8.0), 120 mM NaCl, 0.5% NP-40) supplemented with a protease inhibitor cocktail (Complete, Roche Applied Science), resolved on 10% SDS-PAGE gels and transferred to nitrocellulose membranes (Bio-Rad). Membranes were blocked in TBS with 5% non-fat milk and probed with anti-HIF-1α monoclonal antibody (BD Transduction Laboratories), anti-HIF-2α polyclonal antibody (NB 100-122, Novus), anti-Flag monoclonal antibody (M2, Sigma-Aldrich), mouse monoclonal anti-HA (HA-11, Covance Research Product), anti-EGLN1(PHD2) monoclonal antibody (D31E11, Cell Signaling), anti-GLUT1 polyclonal antibody (NB300-666, Novus Biologicals), mouse monoclonal anti-tubulin (B-512, Sigma-Aldrich) or anti-vinculin monoclonal antibody (Sigma-Aldrich). Bound proteins were detected with horseradish-peroxidase-conjugated secondary antibodies (Pierce) and Immobilon western chemiluminescent horseradish peroxidase substrate (Millipore).

LC-MS. Metabolite levels in samples were determined by negative mode electrospray LC-MS as previously described². Briefly, metabolites were extracted from exponentially growing cells using 80% aqueous methanol (−80 °C) and were profiled by LC-MS. (R)-2HG, 2-oxolutarate and succinate were quantified by

LC-MS in negative mode using multiple reaction monitoring (MRM) on a Quattro micro triple quadrupole mass spectrometer (Waters). Samples were diluted with equal amounts of 25% acetonitrile and 10 µl aliquots were analysed in triplicate. The HPLC column (Luna NH2, 3 µm, 2.0 × 100 mm, Phenomenex) was operated isocratically with 130 mM ammonium acetate pH 5.0 in 37% acetonitrile/water. The MRM transitions were: 117 > 73, 117 > 93 (succinate), 145 > 57, 145 > 101 (2-OG) and 147 > 85, 147 > 129 ((R)-2HG), 0.2 s dwell time for all transitions. Calibration curves were set up with Quant Lynx, using standards dissolved in reaction buffer.

Enzyme activity assays. (R)-2HG (H8378) and (S)-2HG (S765015) were from Sigma. (R)-2HG was free of contaminating 2-OG as determined by LC-MS under conditions that could detect 0.37 µM exogenous 2-OG in 5 mM (R)-2HG (data not shown). Human EGLN1, 2 and 3, collagen P4H-I and FIH, and murine Tet1 and Tet2 were produced in insect cells and purified as described earlier^{34–37}. The plasmids to generate the baculoviruses coding for Tet1 and 2 were a gift from Y. Zhang (University of North Carolina). IC₅₀ values for (R)-2HG and (S)-2HG were determined based on the hydroxylation-coupled stoichiometric release of ¹⁴CO₂ from 2-oxo-[1-¹⁴C]glutarate using synthetic peptides or double stranded oligonucleotides representing the natural targets of the studied enzymes as substrates. These were DLDLEMLAPYIPMDDDFQL (DLD19) for EGLN1, 2 and 3, (PPG)₁₀ for collagen P4H-I, DESGLPQLTSYDCEVNAPIQGSRNLLQGEELRLAL for FIH and 5'-CTATACCTCCTCAACTT(mC)GATCACCGTCTCCGGCG-3' for Tet1 and 2. The K_i values for (S)-2HG for EGLN1, 2 and 3 were determined by adding (S)-2HG in four constant concentrations while varying the concentration of 2-oxo-[1-¹⁴C]glutarate.

To study whether (R)-2HG, which failed to efficiently inhibit EGLN activity, could promote EGLN activity by acting as a cofactor in the place of 2-OG, we determined the amount of 4-hydroxy[³H]proline formed by a specific radiochemical procedure³⁸ using a L-[2,3,4,5-³H]proline-labelled HIF-1α ODDD as a substrate. Collagen P4H-I and a [¹⁴C]-proline-labelled protocollagen³⁹ substrate were used as controls (detecting 4-hydroxy[¹⁴C]proline), and 2-OG (non-labelled) and (S)-2HG were assayed for comparison. The K_m values of EGLN1 and EGLN2 were determined by adding increasing amounts of (R)-2HG while the concentration of the substrate and other cofactors were kept constant.

Generation of a recombinant HIF-1α substrate. The HIF-1α ODDD substrate, spanning residues 356–603 of human HIF-1α, was produced in a BL21(DE3) *E. coli* strain (Novagen) in the presence of L-[2,3,4,5-³H]proline (75 Ci mmol⁻¹, PerkinElmer Life Sciences) and affinity purified in a chelating Sepharose column charged with Ni²⁺ (ProBond, Invitrogen) exploiting a C-terminal His tag⁴⁰. Concentration of the purified substrate was measured by RotiQuant (Carl Roth GmbH) and it was used at K_m concentrations for the distinct EGLN enzymes⁴⁰.

Modelling. EGLN1 active site structure (PDB accession 3HQR²⁹) with N-oxalylglycine (NOG, magnet), Mg²⁺ and a peptide substrate (light blue), was used to model (R)-2HG (green) and (S)-2HG (cyan) into the active site.

Gene expression profiling and gene-set enrichment analysis. Expression data from human brain tumour samples were obtained from The Cancer Genome Atlas (<http://tcga-data.nci.nih.gov/tcga/tcgaHome2.jsp>) and processed as described^{8,28}. In short, 200 expression profiles from glioblastoma multiforme and two non-neoplastic brain samples were generated using three platforms (Agilent 244K, Affymetrix HT-HG-U133A, Affymetrix HuEx), preprocessed using gene centric probe sets and three expression values were integrated through factor analysis. After consensus clustering using 1,740 variably expressed genes, profiles with a negative silhouette metric were identified and removed from the data set, leaving expression profiles from 173 glioblastoma tumour samples. IDH1 mutation status was established for 116 out of the 173 samples.

Eleven gene sets representing response to induced hypoxia were reported in supplementary table 6 from ref. 41. One gene set was from figure 4C in ref. 9. A final set was assembled by one of us (W.G.K.) based on a literature review of genes upregulated by HIF in a wide variety of cell types (*CA9*, *EGLN1*, *EGLN3*, *SLC2A1*, *BNIP3*, *ADM*, *VEGF*, *PDK1*, *LOX*, *PLOD1*, *CXCR4*, *P4HA1*, *ANKRD37*). Single-sample gene-set enrichment analysis (GSEA) was applied as reported previously⁸. Briefly, genes were ranked by their expression values. The empirical cumulative distribution functions (ECDF) of both the genes in the signature as well as the remaining genes were calculated. An enrichment score was obtained by a sum of the difference between a weighted ECDF of the genes in the signature and the ECDF of the remaining genes. This calculation was repeated for all signatures and samples. Z-score transformation was applied to be able to make scores from different gene sets comparable. A positive score indicates gene set activation. A negative value does not indicate inactivation, but rather a lack of effect.

Cell proliferation assays. Cells were plated in 96-well plates (~700 cells per well) with a media change every 3 days. The number of viable cells per well at each time point was measured using an XTT assay (Cell Proliferation Kit II, Roche)

according to the manufacturer's instructions. Spectrophotometrical absorbance at 450 nm was measured 5–6 h after adding the XTT labelling reagent/electron coupling reagent using a microtiter plate reader (Perkin Elmer Life and Analytical Science). For direct cell counting, cells were plated in p60 dishes (~50,000 cells per dish) with a media change every 3 days. The number of viable cells at each time point was measured after trypan blue staining by using an automated cell counter (Invitrogen) according to the manufacturer's instructions.

Soft agar colony formation assay. Approximately 8,000 cells were suspended in a top layer of 0.4% soft agar (SeaPlaque Agarose, BMA products) and plated on a bottom layer of 1% soft agar containing complete DMEM supplemented with 10% FBS in 6-well plates. After 3–4 weeks, colonies were stained with 0.1% iodinitrotetrazolium chloride (Sigma-Aldrich).

Real-time qPCR analysis. Total RNAs were extracted with Trizol reagent (Invitrogen). cDNA synthesis and PCR amplification were performed with Superscript One-Step RT-PCR (Invitrogen) with 2 µg total RNA. *EGLN3* cDNA was amplified with sense primer (5'-GCGTCTCCAAGCGACA-3') and antisense primer (5'-GTCTTCAGTGAGGGCAGA-3'). *VEGF* cDNA was amplified with sense primer 5'-CGAAACCATGAACCTTCTGC-3') and antisense primer 5'-CCTGAGTGGGCACACACTCC-3'). *HIF1A* cDNA was amplified with sense primer (5'-TATTGCACTGCACAGGCCACATTC-3') and antisense primer (5'-TGATGGGTGAGGAATGGGTCACA-3'). *HIF2A* was amplified with sense primer (5'-ACAAGCTCCTCTCCTCAGTTTGCT-3') and antisense primer (5'-ACCCTCCAAGGCTTTCAGGTACAA-3'). *GLUT1* *CAIX* cDNA was amplified with sense primer (5'-TGGAAGAAATCGCTGAGGAAGGCT-3') and antisense primer (5'-AGCACTCAGCATCACTGTCTGGTT-3'). *PDK1* cDNA was amplified with sense primer (5'-ATGATGTCAATCCCAATGGCCC-3') and antisense primer (5'-TGAACATTCTGGCTGGTGACAGGA-3'). As a control, β -actin cDNA was amplified with sense primer (5'-ACCAACTGGGACGACA

TGGAGAAA-3') and antisense primer (5'-TAGCACAGCCTGGATAGCAA CGTA-3')

31. Topaloglu, O., Hurley, P. J., Yildirim, O., Civan, C. I. & Bunz, F. Improved methods for the generation of human gene knockout and knockin cell lines. *Nucleic Acids Res.* **33**, e158 (2005).
32. Kohli, M., Rago, C., Lengauer, C., Kinzler, K. W. & Vogelstein, B. Facile methods for generating human somatic cell gene knockouts using recombinant adeno-associated viruses. *Nucleic Acids Res.* **32**, e3 (2004).
33. Percy, M. J. *et al.* A family with erythrocytosis establishes a role for prolyl hydroxylase domain protein 2 in oxygen homeostasis. *Proc. Natl Acad. Sci. USA* **103**, 654–659 (2006).
34. Hirsila, M. *et al.* Effect of desferrioxamine and metals on the hydroxylases in the oxygen sensing pathway. *FASEB J.* **19**, 1308–1310 (2005).
35. Koivunen, P., Hirsila, M., Gunzler, V., Kivirikko, K. I. & Myllyharju, J. Catalytic properties of the asparaginyl hydroxylase (FIH) in the oxygen sensing pathway are distinct from those of its prolyl 4-hydroxylases. *J. Biol. Chem.* **279**, 9899–9904 (2004).
36. Ito, S. *et al.* Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* **466**, 1129–1133 (2010).
37. Annunen, P. *et al.* Cloning of the human prolyl 4-hydroxylase α subunit isoform α (II) and characterization of the type II enzyme tetramer. The α (I) and α (II) subunits do not form a mixed α (I) α (II) β 2 tetramer. *J. Biol. Chem.* **272**, 17342–17348 (1997).
38. Juva, K. & Prockop, D. J. Modified procedure for the assay of H³- or C¹⁴-labeled hydroxyproline. *Anal. Biochem.* **15**, 77–83 (1966).
39. Kivirikko, K. I. & Myllyla, R. Posttranslational enzymes in the biosynthesis of collagen: intracellular enzymes. *Methods Enzymol.* **82**, 245–304 (1982).
40. Koivunen, P., Hirsila, M., Kivirikko, K. I. & Myllyharju, J. The length of peptide substrates has a marked effect on hydroxylation by the hypoxia-inducible factor prolyl 4-hydroxylases. *J. Biol. Chem.* **281**, 28712–28720 (2006).
41. Benita, Y. *et al.* An integrative genomics approach identifies hypoxia inducible factor-1 (HIF-1)-target genes that form the core response to hypoxia. *Nucleic Acids Res.* **37**, 4587–4602 (2009).

CANCER GENETICS

Evolution after tumour spread

A genetic study of brain cancers in mice and humans reveals distinct mutations in primary tumours and their metastases, suggesting that the two disease ‘compartments’ may require different treatments.

STEVEN C. CLIFFORD

The spread of a primary tumour to secondary sites in the body is a key step in the development of many cancers, and treatment of these secondary metastatic tumours represents one of the foremost challenges in oncology. In an article published on *Nature's* website today, Wu *et al.*¹ describe two new mouse strains that serve as models of metastasis in the childhood brain cancer medulloblastoma. In the mice, primary and metastatic tumours seem to occupy two genetically distinct ‘compartments’, which arise from divergent DNA-sequence mutations that occur after metastasis. The authors also detect similar differences in human medulloblastoma tumours — a finding that may influence the development of anticancer therapies.

Wu *et al.* used an experimental system called Sleeping Beauty mutagenesis^{2,3} to introduce random genetic mutations into cerebellar progenitor cells in the developing brains of two strains of mice. These two new strains were derived by breeding the existing *Tp53^{mut}* and *Ptch^{+/-}* strains, which are predisposed to brain tumours^{1,4}, with a strain that expresses the Sleeping Beauty mutagen in cerebellar progenitors. This system leaves a unique genetic ‘footprint’ at each mutation site, which allows mutated genes to be identified by DNA sequencing. Such mutagenesis experiments are used to identify genes in which mutations frequently arise, because the likelihood of these being involved in tumour development is reasoned to be above average^{2,3}.

As in mouse models of other cancer types^{2,3}, Wu and colleagues’ Sleeping Beauty mutagenesis accelerated the development of medulloblastoma in both mouse strains. The authors identified a range of new and established cancer-related genes that had mutations, including some that have previously been implicated in medulloblastoma. They also observed that, following mutagenesis, mice of both strains developed metastases around a type of brain tissue called the leptomeninges, in patterns that are reminiscent of metastatic human medulloblastoma⁵. The two mouse models thus provided an opportunity to track mutations present in the primary and metastatic disease, and to investigate their genetic provenance.

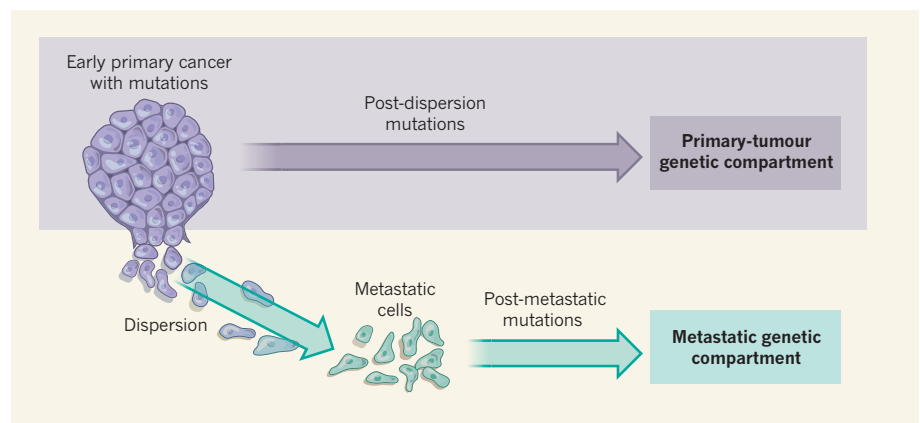


Figure 1 | A bi-compartmental genetic model of cancer metastasis. By analysing the tumours from two strains of mice that model the brain cancer medulloblastoma, Wu *et al.*¹ found differences in the DNA-sequence mutations present in primary and metastatic tumours. They propose that rare cells in the primary tumour that are capable of metastasizing disperse to other sites in the brain, where they form metastases. The cells of the primary tumour and the metastases then continue to accumulate mutations, generating two distinct genetic compartments.

Wu *et al.* found that there were, in general, only a few mutations common to primary and metastatic tumours from the same mouse, but that the mutations in different metastases from the same mouse tended to be more similar to each other. Moreover, certain mutations observed in metastases were detected at only low levels within the primary tumour, and some mutations were unique to one or the other tumour type. The authors conclude that their findings are consistent with a model in which metastases originate from rare cells in the primary tumour, and that, following metastasis, additional mutations accumulate independently — both in the primary tumour (post-dispersion events) and in metastases (post-metastasis events) (Fig. 1).

Turning our attention away from mice, an obvious question is whether primary and metastatic tumours in the human disease also show ‘bi-compartmental’ genetics. Approximately 30% of patients with medulloblastoma already have metastases when they are first diagnosed, and this is associated with a poor prognosis⁵. However, few previous studies have compared the biology of human primary tumours with their associated metastases, mainly because metastases are not routinely biopsied. Despite the limited sample availability, Wu *et al.*¹ show initial evidence of differing

genetics in primary and metastatic tumours from seven human patients.

Further investigation is required to establish whether the authors’ findings are broadly relevant to human medulloblastoma. The human disease exhibits⁵ more complex patterns of metastases than are observed in mice, and is classified into four molecular subgroups (WNT, SHH, Group 3 and Group 4), which each display distinct biological and clinical characteristics⁶. The *Ptch^{+/-}* mice used by Wu *et al.*¹ develop SHH-associated medulloblastomas⁴; similar mutagenesis-driven approaches using existing mouse models of other medulloblastoma disease groups, such as WNT⁷, might prove informative.

Perhaps the most urgent question arising from this study¹ is whether the genetic differences between the two disease compartments lead to distinct biological features that make them respond differently to treatment. In mice, these compartments remain genomically characterized entities, the biological and therapeutic importance of which is untested. In humans, clinical-trial data show⁵ that primary and metastatic sites respond similarly to current therapies (with cure achieved at both sites) in around 60% of children with metastatic disease, but a more objective assessment of treatment response is confounded by the fact that

primary tumours are mostly removed by surgery prior to treatment. Wu *et al.* provide initial evidence that the tumour compartments may respond differently to current therapies in certain patients, but they rightly caution that these effects could also relate to clinical factors such as radiotherapy being delivered at different intensities to different tumour sites.

Some of the mutations identified by Wu and colleagues' experiments may also reveal biological processes or pathways that could offer drug targets for the improved treatment of primary tumours, metastases, or both. The new mouse strains provide excellent models in which to test this possibility. The multitude and variety of mutations described by Wu *et al.*¹ are noteworthy, but the next challenge is to determine which of them can drive tumour development, which are therapeutically relevant, and which occur at sufficient

frequency in the human disease to warrant their pursuit as potential targets. The authors justifiably reason that targets that are common to primary tumours and metastases, in both humans and mice, are those most attractive for further development. However, only one cellular pathway, insulin-dependent signalling, meets these criteria on the basis of their current data.

Providing answers to all these questions will require further biological investigation across species, as well as clinical studies. An additional challenge is posed by the fact that there are fewer than 700 cases of medulloblastoma per year in Europe. More routine biopsy and characterization of human metastases will be essential, and the impetus and ethical justification for such a fundamental change to clinical practice will, at least in part, come from experimental studies such as those

presented here. Time will tell whether this tale of Sleeping Beauty and mice develops into a clinically relevant human paradigm. ■

Steven C. Clifford is at the Northern Institute for Cancer Research, Newcastle University, Newcastle upon Tyne NE1 4LP, UK.
e-mail: steve.clifford@ncl.ac.uk

1. Wu, X. *et al.* *Nature* <http://dx.doi.org/nature10825> (2012).
2. Collier, L. S., Carlson, C. M., Ravimohan, S., Dupuy, A. J. & Largaespada, D. A. *Nature* **436**, 272–276 (2005).
3. Copeland, N. G. & Jenkins, N. A. *Nature Rev. Cancer* **10**, 696–706 (2010).
4. Goodrich, L. V., Milenković, L., Higgins, K. M. & Scott, M. P. *Science* **277**, 1109–1113 (1997).
5. Pizer, B. L. & Clifford, S. C. *Br. J. Neurosurg.* **23**, 364–375 (2009).
6. Taylor, M. D. *et al.* *Acta Neuropathol.* <http://dx.doi.org/10.1007/s00401-011-0922-z> (2011).
7. Gibson, P. *et al.* *Nature* **468**, 1095–1099 (2010).